The Astrophysical Journal, 892:1 (27pp), 2020 March 20 https://doi.org/10.3847/1538-4357/ab7914

The Regulation of Galaxy Growth along the Size–Mass Relation by Star Formation, as Traced by Hα in KMOS^{3D} Galaxies at $0.7 \lesssim z \lesssim 2.7$*

David J. Wilman[1,2], Matteo Fossati[1,2,3], J. Trevor Mendel[2,4,5], Roberto Saglia[1,2], Emily Wisnioski[2,4,5], Stijn Wuyts[2,6], Natascha Förster Schreiber[2], Alessandra Beifiori[1,2], Ralf Bender[1,2], Sirio Belli[2], Hannah Übler[2], Philipp Lang[7,2], Jeffrey C. C. Chan[8,2], Rebecca L. Davies[2], Erica J. Nelson[2,9], Reinhard Genzel[2], Linda J. Tacconi[2], Audrey Galametz[2,10], Richard I. Davies[2], Dieter Lutz[2], Sedona Price[2], Andreas Burkert[1,2], Ken-ichi Tadaki[2,11], Rodrigo Herrera-Camus[2,12], Gabriel Brammer[13], Ivelina Momcheva[14], and Pieter van Dokkum[15]

[1] Universitäts-Sternwarte München, Scheinerstrasse 1, D-81679 München, Germany; matteo.fossati@durham.ac.uk
[2] Max-Planck-Insitut für extraterrestrische Physik, Giessenbachstrasse, D-85748 Garching, Germany
[3] Centre for Extragalactic Astronomy and Institute for Computational Cosmology, Durham University, South Road, Durham, DH1 3LE, UK
[4] Research School of Astronomy and Astrophysics, Australian National University, Canberra, ACT 2611, Australia
[5] 3ARC Centre of Excellence for All Sky Astrophysics in 3 Dimensions (ASTRO 3D), Australia
[6] Department of Physics, University of Bath, Claverton Down, Bath, BA2 7AY, UK
[7] Max Planck Institute for Astronomy (MPIA), Königstuhl 17, D-69117, Heidelberg, Germany
[8] Department of Physics and Astronomy, University of California Riverside, 900 University Avenue, Riverside, CA 92521, USA
[9] Harvard-Smithsonian Center for Astrophysics, 60 Garden St., Cambridge, MA 02138, USA
[10] Observatoire de Genève, Université de Genève, 51 Ch. des Maillettes, CH-1290 Versoix, Switzerland
[11] National Astronomical Observatory of Japan, 2-21-1 Osawa, Mitaka, Tokyo 181-8588, Japan
[12] Astronomy Department, Universidad de Concepción, Barrio Universitario, Concepción, Chile
[13] Cosmic Dawn Center, Niels Bohr Institute, University of Copenhagen, Juliane Maries Vej 30, DK-2100 Copenhagen, Denmark
[14] Space Telescope Science Institute, Baltimore, MD 21218, USA
[15] Astronomy Department, Yale University, 52 Hillhouse Avenue, New Haven, CT 06511, USA
Received 2019 April 29; revised 2020 February 20; accepted 2020 February 20; published 2020 March 19

Abstract

We present half-light sizes measured from Hα emission tracing star formation in 281 star-forming galaxies from the KMOS^{3D} survey at $0.7 \lesssim z \lesssim 2.7$. Sizes are derived by fitting 2D exponential disk models, with bootstrap errors averaging 20%. Hα sizes are a median (mean) of 1.19 (1.26) times larger than those of the stellar continuum—which, due to radial dust gradients, places an upper limit on the growth in stellar size via star formation—with just ∼43% intrinsic scatter. At fixed continuum size the Hα size shows no residual trend with stellar mass, star formation rate, redshift, or morphology. The only significant residual trend is with the excess obscuration of Hα by dust, at fixed continuum obscuration. The scatter in continuum size at fixed stellar mass is likely driven by the scatter in halo spin parameters. The stability of the ratio of Hα size to continuum size demonstrates a high degree of stability in halo spin and in the transfer of angular momentum to the disk over a wide range of physical conditions and cosmic time. This may require local regulation by feedback processes. The implication of our results, as we demonstrate using a toy model, is that our upper limit on star-formation-driven growth is sufficient only to evolve star-forming galaxies approximately *along* the observed size–mass relation, consistent with the size growth of galaxies at constant cumulative comoving number density. To explain the observed evolution of the size–mass relation of star-forming disk galaxies, other processes, such as the preferential quenching of compact galaxies or galaxy mergers, may be required.

Unified Astronomy Thesaurus concepts: Galaxy evolution (594); Galaxy structure (622); Galaxy disks (589); Star formation (1569); Spectroscopy (1558)

Supporting material: machine-readable

1. Introduction

Most star-forming galaxies in the universe above stellar masses of $M_* \sim 10^9\,M_\odot$ have most of their stars in disks (e.g., Wuyts et al. 2011; van der Wel et al. 2014b). These are stable, rotationally supported structures that, in the absence of dramatic events such as major mergers, survive at least for the Hubble time with disk galaxies still dominant in the local star-forming population. Galaxy disks typically have radial surface brightness profiles well described by a declining exponential function (although an improved description of many local stellar disks is a broken exponential law, becoming either steeper or more shallow beyond a break radius; Erwin et al. 2008). Disks exist not only in the stellar component but also in the gas that feeds them, and although the thickness of stellar disks and turbulence in gas disks can vary with time, basic disk structures with dominant rotational support exist to high redshifts (e.g., Förster Schreiber et al. 2006, 2009; Genzel et al. 2006; Kassin et al. 2012; Livermore et al. 2015), up to at least $z \sim 3$ (Turner et al. 2017), and are dominant among the high-mass population by $z \sim 2.2$ (Wisnioski et al. 2015; Stott et al. 2016), with evidence that they are common even in the most compact (Wisnioski et al. 2018) and passively evolving old galaxies at that redshift (e.g., McGrath et al. 2008; van der Wel et al. 2011; Chang et al. 2013; Newman et al. 2015; Toft et al. 2017; Hill et al. 2019).

The structure of massive star-forming galaxies is made not only of rotating disks but also by central dispersion-dominated bulges (Lang et al. 2014). These could be the result of violent star formation from low angular momentum cold gas in the

* Based on observations obtained at the Very Large Telescope (VLT) of the European Southern Observatory (ESO), Paranal, Chile (ESO program IDs 092.A–0091, 093.A–0079, 094.A–0217, 095.A–0047, 096.A–0025, 097.A–0028, 098.A–0045, 099.A–0013).

center of galaxies, although they can also form during merger events. Submillimeter observations reveal very high rates of highly obscured star formation at the center of massive galaxies at high redshift (Tadaki et al. 2017). Such events appear to coexist with more extended and less obscured star formation (e.g., Chen et al. 2017), such that star-forming disks tend to retain an exponential profile, even in the presence of a bulge or bar.

Observations in the local universe indicate that stars form predominantly from the dense and cool molecular gas component, with star formation surface density well correlated to the molecular gas surface density (Bigiel et al. 2008), with a slope close to unity in the disk regime, implying a constant timescale for the depletion of molecular gas by star formation. Interestingly, the star formation also appears to track the existing stars, to first order. For example, there exists a relation between the local density of star formation and that of stars (González Delgado et al. 2016). This reflects on the spatial extent of these components, such that in terms of half-light sizes, the sizes of the star-forming disks are found to be extremely similar to that of the stellar disk in the local universe (Fossati et al. 2013).

As gas accretes onto a galaxy, it still carries much of the angular momentum from the cosmic filaments that feed the galaxy and its halo (Fall & Romanowsky 2018). Smooth accretion of gas with a consistent axis of angular momentum leads to the formation of gas disks. While the mean specific angular momentum of disks is similar to that of their halo (Burkert et al. 2016), the distribution within any single galaxy of angular momentum from newly accreted halo gas is expected to extend to both lower and higher values than found in typical galaxy disks (e.g., Dalcanton et al. 1997; van den Bosch 2001; Dutton 2009). The high angular momentum material can be transported to large radii, where it will exist in a diffuse atomic or ionized component unable to form new stars, while the low angular momentum material can be removed in energetic supernova-driven winds. Such winds are particularly effective at removing material from low-mass and compact galaxies (Dutton 2009) but can delay the evolution of higher-mass galaxies via high-redshift ejection and reincorporation (e.g., Hirschmann et al. 2013).

The existence at $0 < z \lesssim 3$ of a main sequence (MS) of star formation, relating the galaxy star formation rate (SFR) to the stellar mass with a small scatter (∼0.3 dex; e.g., Noeske et al. 2007; Whitaker et al. 2014; Gavazzi et al. 2015; Schreiber et al. 2015), implies that the combined processes of gas accretion and star formation must be smooth and stable over the relatively short timescales to which we are sensitive with typical star formation indicators. Moreover, these small variations in SFR at fixed stellar mass seem to have no measurable dependence on the galaxy size, but are rather driven by the molecular content of galaxies or its depletion rate (Saintonge et al. 2011; Tacconi et al. 2018). The mass of the cold gas reservoir is also the main driver of the cosmic evolution of the star formation activity, with gas-rich galaxies at $z \sim 1$–2, forming stars much more rapidly than in the local universe (Madau & Dickinson 2014; Whitaker et al. 2014).

The relationship between the local density of star formation and that of stellar mass found in the local universe appears to extend to at least $z \sim 1$ (Wuyts et al. 2013). Half-light sizes in the Hα emission line tracing unobscured star formation are similar to or slightly larger than the size in continuum light in both individual highly star-forming galaxies (Nelson et al. 2012) and the stacked averages for normally star-forming galaxies (Nelson et al. 2016a). Using 3D-*HST* slitless spectroscopic data (van Dokkum et al. 2011; Brammer et al. 2012; Skelton et al. 2014; Momcheva et al. 2016), Nelson et al. (2016a) show that the stacked average Hα profiles of star-forming galaxies with higher- or lower-than-normal SFR for their stellar mass are self-similar, changing only in normalization and not half-light size. The sizes of molecular gas disks themselves are not easily measured at high redshift in normally star-forming galaxies. Where measured, they appear similar in extent to the stellar or star-forming disks (Tacconi et al. 2013; Bolatto et al. 2015), while in the highly star-forming, high-mass population the situation is more complex: highly compact dust emission can coexist with more extended emission from tracers of molecular gas such as CO (Calistro Rivera et al. 2018).

KMOS^{3D} is a unique 75-night guaranteed time program with the ESO Very Large Telescope (VLT) with the second-generation instrument KMOS (*K*-band Multi-Object Spectrograph; Sharples et al. 2012) targeting the Hα+[N II] emission-line complex in ∼740 galaxies selected to have a magnitude $Ks < 23$ and in the range $0.7 \lesssim z \lesssim 2.7$ (Wisnioski et al. 2019, hereafter W19; Wisnioski et al. 2015). The multiplexing capabilities of KMOS allow us to target more galaxies and with deeper observations than was possible with single-object integral field units (IFUs) such as SINFONI (e.g., the SINS survey; Förster Schreiber et al. 2009), complimenting contemporary work on smaller numbers of objects featuring the high spatial resolution available with adaptive optics (e.g., Förster Schreiber et al. 2018).

In this paper we use KMOS^{3D} data to map Hα and measure Hα disk sizes in individual star-forming galaxies across a wide range in redshift and SFR. We examine whether the stacked results of Nelson et al. (2016a) apply for individual galaxies and whether size growth via star formation is correlated with the stellar mass and SFR or is driven by other fundamental parameters across a wide baseline in redshift, including the peak of the cosmic star formation activity. KMOS^{3D} offers several advantages compared to 3D-*HST* for a study of this nature: it is significantly deeper, its spectral resolution allows us to resolve the Hα+[N II] emission-line complex, and observations in the *YJ* to *Ks* band allow us to trace Hα emission over a larger redshift range. Being a seeing-limited ground-based survey, this goes at the expense of spatial resolution.

Following our brief introduction to the KMOS^{3D} survey in Section 2, Sections 3–5 give a detailed account of how we go from raw KMOS data to accurate size measurements of KMOS^{3D} galaxies with well-calibrated errors. Readers primarily interested in our results on—and interpretation of—the size growth of star-forming galaxies may skip to Section 6. For readers interested in the technical steps, we describe the basic data reduction in Section 3 and the generation of Hα maps and profiles in Section 4. In Section 5 we describe the flagging procedures used to verify our sample and show that it is not biased with respect to normally star-forming MS galaxies. We also release to the community the size measurements derived in this work. We then examine which parameters control the Hα size of KMOS^{3D} galaxies in Section 6, and in Section 7 we discuss what this means for our understanding of how galaxies grow in size through star formation. Our key conclusions are presented in Section 8. Throughout this paper we assume a flat

ΛCDM cosmology with $H_0 = 70\ \mathrm{km\ s^{-1}\ Mpc^{-1}}$, $\Omega_m = 0.3$, and $\Omega_\Lambda = 0.7$.

2. The KMOS^{3D} Survey

KMOS^{3D} takes advantage of the unique multiplexing and spatially resolved near-infrared (NIR) spectroscopic capabilities of KMOS, as well as the large collecting area of the 8.2 m VLT mirror, targeting the Hα+[N II] emission-line complex in galaxies at $0.7 \lesssim z \lesssim 2.7$. This provides simultaneous flux and kinematic maps of the ionized gas for up to 24 galaxies in one exposure by deploying 24 configurable arms in the 7.′2 field of view (FOV), each hosting a $2.''8 \times 2.''8$ IFU.

The first year of the KMOS^{3D} survey was described by Wisnioski et al. (2015). KMOS^{3D} targets galaxies selected from the 3D-*HST* grism (Brammer et al. 2012; Skelton et al. 2014; Momcheva et al. 2016) and CANDELS imaging (Grogin et al. 2011; Koekemoer et al. 2011) surveys with the *Hubble Space Telescope* (*HST*) in the COSMOS, GOODS-South, and UDS deep fields accessible from Paranal. Targets are selected to have a magnitude $Ks < 23$ and a known spectroscopic or grism redshift (grism redshifts from 3D-*HST* have an accuracy of ≈1000 km s^{-1}; Momcheva et al. 2016; Fossati et al. 2017) for which the spectrum around Hα should be relatively free of atmospheric OH lines, and visible in the KMOS *YJ*, *H*, or *K* bands. We apply no prior selection on SFR or Hα flux in order to avoid selection bias and sample the full range of galaxies down to our detection limits. Due to the unique multiplexing capabilities of KMOS, we are able to observe galaxies from ∼3 to 30 hr by retargeting objects with weak detections to improve the signal-to-noise ratio (S/N).

Observations for KMOS^{3D} were carried out from 2013 October until 2018 April following an object-sky-object observation pattern such that each object exposure is adjacent to a sky exposure in the same IFU. Three IFUs were placed on stars to trace the variable spatial point-spread function (PSF) and throughput from exposure to exposure, leading to simultaneous observations of up to 21 galaxies per exposure.

In this work we consider all data taken up until 2017 April, comprising 645 galaxies targeted for observations of Hα and [N II]. Data were taken in a range of observing conditions, with PSF minor-axis FWHM ranging between 0.″3 and 0.″92 and a median of 0.″456. The final KMOS3D data set is fully described in W19. SFRs used in this paper are computed using the data and method described by Wuyts et al. (2011), based on infrared, UV, and optical observations and thus independent of our Hα measurements.

3. Data Reduction

Our reductions in this paper are intermediate between the early data reduction described by Wisnioski et al. (2015) and that described in the data release paper (W19). We refer to W19 for much of the reduction procedure, noting where implementation of specific steps differs. In particular, we describe in detail the steps that optimize the background subtraction and astrometry, and which were tailored to allow a robust extraction of Hα profiles and sizes of galaxies.

3.1. Basic Reduction

All our basic calibration steps, with the exception of the sky and background subtraction, are identical to those described by W19. We make use of the Software Package for Astronomical Reductions with KMOS (SPARK) code, which works within the ESO pipeline execution tool (ESOREX), supplemented with some custom tools written in the IDL and PYTHON languages. This includes masking of bad pixels and flattening at the detector level; reconstruction of data cubes, including a refined wavelength calibration using skylines and a heliocentric correction; followed by a correction for the spatial illumination uniformity, and flux calibration using standard-star observations. During this last step, we used the flux from stars observed in the same setup as the galaxies to correct for frame-to-frame variations in the throughput. Bad frames are inspected and removed. Skyline subtraction was applied using the standard method in SPARK, which subtracts an adjacent sky frame with skylines scaled to optimally match the science observation (Davies 2007).

Once individual frames are generated, it is essential to subtract a residual background level per frame: not doing so results in a factor of three reduction in continuum S/N in the final co-adds, primarily due to the significant variations in instrumental, sky (e.g., twilight and moon illumination), and thermal (especially in *K* band) background between object and adjacent sky frames. Instrumental variations include a readout-channel-dependent effect, which can vary frame to frame. To account for this effect, we derive and subtract a background value for each of the readout channels of the detectors.

3.2. Astrometric Registration, Improved Background Subtraction, and Generation of Combined Cubes

After the reconstruction of individual frames, *Partial* combined cubes, defined to be the co-add of the data taken for a given galaxy within a given observing setup (commonly one per observing run), are generated assuming astrometric shifts between frames equal to the average of the measured shifts for the three stars included in the same setup (this accounts for the telescope dithering and the gradual drift of the KMOS arm positions).

We also generate 100 bootstrap cubes obtained by randomly resampling the input frames for each partial combine. We use these cubes for the propagation of uncertainties. In this work, we make use of a modified combined noise cube with the aim to obtain a robust estimate of the spectral uncertainty close to the edges of the cubes, where few exposures are available (given that SPARK estimates variance from the distribution of values in each exposure). We derive a single variance spectrum per cube using the SPARK variance estimate in spaxels to which at least 75% of the total number of exposures have contributed. We then scale this spectrum by the exposure time of each pixel in the cube.

To achieve the best S/N and image quality for our data in the final cubes, we further process the individual frames to obtain a flat background and an accurate registration of the astrometry between frames observed in different runs. To do this, we generate images of each galaxy by collapsing the KMOS partial combine data cubes along the wavelength axis. The galaxy continuum is well detected for most sources brighter than our $Ks = 23$ magnitude limit, with an increasing fraction of nondetections in continuum close to this limit (∼11% in the range $Ks = 22.5 - 23$).

At this stage, the partial combine cubes retain a residual, negative background caused by the overestimation of background levels in individual frames due to the contribution of the source. Its magnitude in bright sources is ≲10% of the

variation in background level between exposures and decreases in fainter sources, but it is systematic and limits the depth of our final mosaics. Thus, we derive an additive correction to the background as described below.

We first convolve the *HST* image, selected in the nearest available band (WCF3 F125W for KMOS *YJ*, WFC3 F160W for KMOS *H* and *K*) with a multi-Gaussian kernel to optimally convolve the *HST* PSF to that of our KMOS PSF image for that galaxy. Each model solution is defined by an astrometric offset, a normalizing flux scale factor, and an additive background correction per readout channel that contributes in the partially combined datacube. Each model image is generated by projecting the convolved *HST* image onto the KMOS pixel grid, cropping to the KMOS FOV, and then adding the background correction image. We use the MPFIT nonlinear least-squares fitting algorithm to find the minimum chi-squared solution (Markwardt 2009).[16] To ensure that we do not get stuck in a local minimum, we iterate over the initial guess for the astrometric centroid on a grid of 1 pixel resolution, allowing centroids within ±30% of the FOV from the image center, selecting the solution that gives us the global minimum chi-squared. The full procedure is repeated for each bootstrap cube (except that the initial guess for the astrometric centroid is now fixed) to help evaluate errors and degeneracy in the astrometric registration.

Of 166 objects with multiple setups, the median residual shift is ∼1.33 KMOS pixels (∼0.″27), with ∼27% of shifts above 2 KMOS pixels (0.″4), ranging as high as 4.35 pixels (0.″87).[17] Not accounting for such shifts artificially blurs the galaxy by an average of ∼2 kpc and up to ∼7 kpc. These shifts are caused by the variations in the calibration parameters of individual KMOS arms, which are periodically tweaked by the observatory to ensure that the arm positioning remains within specifications. Fits to the individual partial combines are visually inspected, and a new list of astrometric shifts is derived by combining the frame-to-frame shifts measured using PSF stars and the setup-to-setup shifts from the fits of partial combines. We also subtract the best-fit background image from each individual exposure contributing to a given partial combine cube.

With the updated list of astrometric shifts, we combine all frames contributing to a single object and regenerate the bootstrap cubes. This produces our final *total* combined data cubes. At this stage we also derive the instrumental resolution for each cube and its associated PSF image. For a detailed description of these procedures we refer the reader to W19. No correction is applied during the fit of partial combines to the absolute astrometry. This is done by fitting the total combines in order to have the deepest KMOS images register onto the *HST* astrometry. This last fit does not include a background level, as the background has already been flattened during the previous step.

Each astrometric solution is now visibly inspected by looking at the object centroid in the collapsed KMOS image and in the model *HST* image. In 615 of 645 cases the automated solution is good, as in the example case shown in Figure 1. These imply a median shift of ∼1 KMOS pixel (0.″2) with a tail extending to >5 pixels (>1″) and a median bootstrap error of

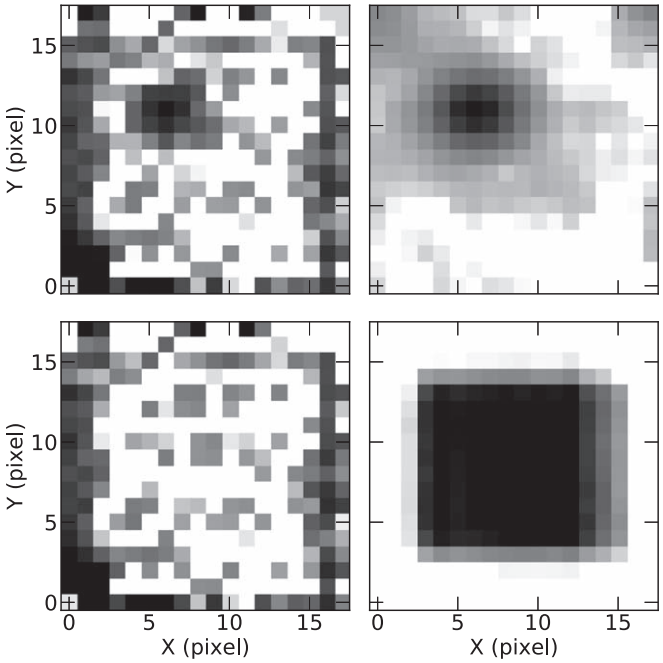

Figure 1. Example of astrometric registration fitting procedure. The collapsed KMOS continuum image (top left) is fit to the PSF-convolved and resampled CANDELS image in the nearest band, including a shift (top right), leaving the residual image (bottom left). Fitting weights are applied (bottom right), which down-weight the outer part of the image (subject to cosmetic effects and lower total exposure). For this example case (COS4_12148) a large shift of 2.5 pixels each of *X* and *Y* is required.

∼0.1 pixels, with errors up to ∼1 pixel. This is consistent with the shifts computed for the same object observed in multiple setups and with the expected positioning accuracy of the KMOS arms. In 26 cases we apply a manual shift, of which for 15 it was necessary to inspect also the Hα image. We used this image to confirm the low-S/N continuum center, finding consistency with the Hα image in all cases. Only four targets are not visible in continuum or Hα; for these no astrometric correction is applied. Manual astrometric solutions are usually accurate within a pixel, with the exception of some low-S/N continuum or Hα-based centroids that can be less accurate (up to ∼2 pixels).

The astrometric correction derived above is applied, registering data cubes and bootstrap cubes to the *HST* astrometry. These astrometrically and background-corrected cubes are considered our final datacube products.

4. Generation of Maps and Profiles

4.1. Emission-line Measurement

To fit the Hα+[N II] emission-line complex, we utilize our versatile IDL-based emission-line fitting software KUBEVIZ[18] (see, e.g., Fumagalli et al. 2014; Fossati et al. 2016). KUBEVIZ can be operated in interactive or batch modes, and provides the user with full access to the options provided by MPFIT, which fits the continuum and emission lines. Gaussian fits to emission lines automatically account for the (known) spectral resolution of the instrument as a function of wavelength.

[16] https://www.physics.wisc.edu/~craigm/idl/fitting.html

[17] W19 reports a lower fraction of large shifts because in that paper we divided the list of exposures into smaller units with smaller shifts for the final data release.

[18] KUBEVIZ is made publicly available at https://github.com/matteofox/kubeviz, and http://www.mpe.mpg.de/~dwilman/kubeviz and is easily adaptable to new instruments.

We derived accurate maps of emission-line flux down to low surface brightness levels, as well as velocity maps and masks of good kinematic fits as described below.

4.1.1. Kinematic Fits

Our first fit is mostly aimed at generating velocity and dispersion maps. To improve the S/N per pixel, the flux, noise, and bootstrap cubes are median-smoothed along spatial axes with a top-hat smoothing kernel of 3 × 3 spaxels. We fit the spectral continuum underlying the Hα+[N II] emission-line complex, assuming a constant value, independently computed for each spaxel, $C_{x,y}$. This is the inverse-variance weighted average value within spectral continuum windows defined to either side of the Hα line, corresponding to between 2000 and 5000 km s^{-1} in rest-frame velocity offset (thus excluding [N II] and [S II] lines). The spectral region from 12680 to 12710 Å, containing the strongest part of the atmospheric O2 feature, and regions within 10 Å of either end of the spectrum are excluded. We generate continuum-subtracted cubes $CS_{x,y,\lambda}$ by subtracting the continuum for all spaxels from the flux cube, $F_{x,y,\lambda}$, in symbols:

$$CS_{x,y,\lambda} = F_{x,y,\lambda} - C_{x,y}. \quad (1)$$

We simultaneously derive kinematic and flux information for the Hα, [N II] λ6583, and [N II] λ6548 emission lines by fitting the inverse-variance weighted continuum-subtracted spectra for each spaxel. We fit a single Gaussian line profile for each emission line, which accounts for the redshifting and instrumental line broadening of the specific KMOS observation, returning the rest-frame velocity and intrinsic dispersion of the ionized gas. Since, for this step, we are interested in robustly detected emission lines, these fits are constrained such that lines have a minimum of zero flux. Multiple lines all share a single velocity and dispersion, and the ratio of flux in the two [N II] lines is fixed to the value from atomic physics (3.071; Storey & Zeippen 2000). As a result, output 2D maps are generated for each fit parameter (line flux, velocity, dispersion).

The full continuum and emission-line fitting process is repeated for each bootstrap cube, generating 100 bootstrap realizations of the fitting parameters. These are used to generate images where each spaxel represents the probability of nonzero Hα line flux $P_{f_{\rm H\alpha}>0}$ and positive nonzero dispersion $P_{\sigma>0}$. $P_{f_{\rm H\alpha}>0}$ represents a detection significance of the flux per spaxel, and $P_{\sigma>0}$ that the emission line is significantly resolved. High values $P_{\sigma>0} \gtrsim 0.9$ provide a good indication that MPFIT has picked up a real feature in the spectrum, rather than a noise or skyline residual spike, and correlates well with regions where the velocity map is relatively smooth. Intrinsic velocity dispersions of $\sigma \gtrsim 25$ km s^{-1} are usually well resolved.

4.1.2. Masking

We then generate an automated spaxel mask to identify spaxels with trustworthy kinematic fits. Within this mask, a good (unmasked) spaxel must meet the following conditions:

$$(f_{\rm H\alpha} > 0.) \quad \text{and} \; (0 < \sigma \leqslant \sigma_{\rm max}) \; \text{and}$$
$$(P_{f_{\rm H\alpha}>0} \geqslant 0.95) \; \text{and} \quad (P_{\sigma>0} \geqslant 0.9), \quad (2)$$

where we set $\sigma_{\rm max} = 250$ km s^{-1} to exclude broad line features, given that our primary goal at this stage is to define the velocity map. This mask is applied to the velocity map of example galaxy U4_25642 in the left panel of Figure 2.

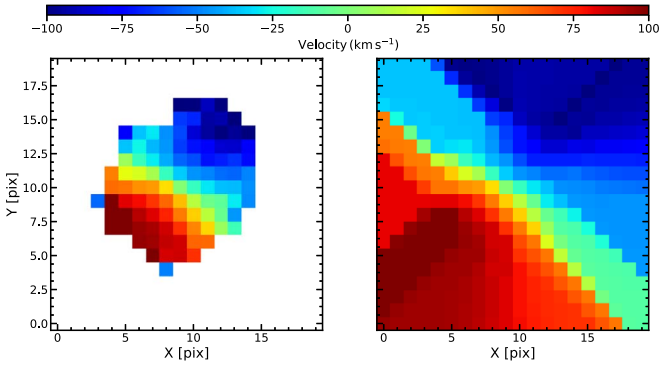

Figure 2. Example masked velocity field (left) and grown velocity map (right) for galaxy U4_25642. Blank pixels in the left panel are masked, using the pre-sigma-clipping mask discussed in the text. Additional automated sigma clipping and (in some cases) additional manual editing of the mask are then applied, followed by smoothing and growth by extrapolation of the masked velocity map to generate the grown velocity map as described in the text.

We then apply some further steps designed to throw out potential outlier spaxels in the velocity map. The sigma-clipped mean and rms velocities are computed, and any spaxel with velocity outside the range mean ± 3 × rms is thrown out. Isolated unmasked spaxels are removed, and the remaining are smoothed with a 3 × 3 top-hat filter. This leaves us with a conservative mask and a smooth velocity map. While conservative, our maps are consistent with those presented in other KMOS^{3D} papers focused on galaxy kinematics (see, e.g., Wisnioski et al. 2015).

Galaxies with <3 valid spaxels are dropped, as are poor fits established by inspecting objects with <10 valid spaxels. This relatively small minimum number of spaxels is sufficient to establish a zero-point and lack of chaotic variation in the velocity maps of compact galaxies, which is sufficient for our purpose. This leaves 462 galaxies. All velocity maps are visually inspected, resulting in the removal of a further seven cases (leaving 455 galaxies) and the manual correction of 117 masks. These fix cases in which the rotation curve gets truncated by the sigma-clipping procedure, or there are systematic fits to sky features that are usually spatially offset from the galaxy and are not sigma-clipped.

We then extrapolate the rest-frame velocity map to the edges of the KMOS FOV to generate a *grown velocity map*, $dv_{\rm rest}(x, y)$. This involves setting the velocity of masked spaxels to the average value of their neighbors, starting with those with unmasked neighbors and iterating up until the point at which the whole area is filled. While this extrapolation cannot pick up changes in the rotation curve in the outer, low surface brightness parts of the galaxy, it involves minimal assumptions. This procedure is repeated for each bootstrap iteration independently. The right panel of Figure 2 shows the resultant grown velocity map for U4_25642.

4.1.3. Deep Emission-line Flux Maps

To derive our final Hα flux maps, we reapply continuum fitting and subtraction, this time to the *unsmoothed* cubes. Then, we integrate the flux within each spaxel of the continuum-subtracted cube, centered at the wavelength specified for Hα, assuming the "grown" velocity map. In the observed frame this is $\lambda_{\rm cen}(x, y) = \lambda_{\rm H\alpha}.(1 + dv_{\rm rest}(x, y)/c) \times (1 + z)$ with a

window width of $\pm 200 \times (1+z)$ km s^{-1}. This window is sufficiently wide to encompass offsets from the true velocity caused by the extrapolation process into the outer disk (see, e.g., Lang et al. 2017), without losing too much flux in the wings of all but the broadest emission lines. A broader window would lead to reduced S/N. This narrowband extraction generates a map of Hα flux down to regions of low S/N, where parametric fitting fails or becomes unreliable. Our window-integrated Hα map is computed as

$$f_{\mathrm{H}\alpha,\mathrm{WIN}}(x,y) = \Delta\lambda \cdot \Sigma_{\lambda_{\mathrm{upper}}(x,y)}^{\lambda_{\mathrm{lower}}(x,y)} CS_{x,y,\lambda}, \quad (3)$$

with the bounds in the sum given by $\lambda_{\mathrm{cen}}(x,y) \pm (200\ \mathrm{km\ s^{-1}\ c^{-1}}) \times (1+\mathrm{z})$ and $\Delta\lambda$ being the spectral step (in Å) of the data cubes. This procedure is again repeated for each of the 100 bootstrap realizations.

Finally, we correct these flux maps $f_{\mathrm{H}\alpha,\mathrm{WIN}}(x,y)$ for flux lost outside the ± 200 km s^{-1} window from the tails of broad lines. An alternate mask is used to define regions for which the velocity and dispersion from Gaussian fitting are usable for this purpose: this is equivalent to Equation (2), but with a relaxed upper limit on dispersion σ: in this case $\sigma_{\mathrm{max}} = 1000$ km s^{-1} for $\mathrm{S/N}_{\mathrm{H}\alpha} \geqslant 4$ and $\sigma_{\mathrm{max}} = 400$ km s^{-1} for lower $\mathrm{S/N}_{\mathrm{H}\alpha} < 4$ (thresholds selected from visual inspection of fits to broad lines). Isolated, unmasked spaxels are thrown out, and manual edits made to our earlier masks are reapplied to this mask. The dispersion measured by the fit is used to compute the fraction of emission flux that falls outside our ± 200 km s^{-1} window, and this correction is applied to the flux map:

$$f_{\mathrm{H}\alpha,\mathrm{WINcor}}(x,y) = f_{\mathrm{H}\alpha,\mathrm{WIN}}(x,y)/c_{\sigma_{200}}, \quad (4)$$

with $c_{\sigma_{200}}$ the two-tailed cumulative distribution function for a Gaussian, evaluated at $\sigma_{200} = \sigma/200$ km s^{-1}. The correction is usually small except in regions containing very broad lines ($c_{\sigma_{200}} = 0.955$, 0.576, and 0.197 for velocity dispersions $\sigma = 100$, 250, and 1000 km s^{-1}, respectively). Outside the mask we do not have reliable data to make a correction. However, these regions usually correspond to the outer and low surface brightness parts of galaxies, where the dispersions are typically low ($\lesssim$100 km s^{-1}; Wisnioski et al. 2015), and so any correction would be very small.

This correction is also applied to the bootstrap realizations, and we consider $f_{\mathrm{H}\alpha,\mathrm{WINcor}}(x,y)$ our best estimate of the flux map for the Hα emission line, from now on simply known as the Hα flux map or image.

4.2. Image Fitting in 2D

We now model the radial surface brightness profiles by fitting the images of our galaxies in 2D using the Levenburg–Marquardt solver from the image fitting code IMFIT[19] (Erwin 2015). Our goal is to quantify the distribution of continuum flux (in particular, the F160W band tracing older stars) and Hα flux (tracing ongoing star formation) obtained with the spectral windows method defined above. We fit continuum profiles with a Sérsic profile (Sérsic 1968) and Hα profiles with a pure exponential, motivated by the results of Nelson et al. (2016a). Despite radial and/or azimuthal variations from a pure exponential profile, the best-fit solution will account for the mean surface brightness at fixed radius, and as such it provides a measure of the average star formation at that radius, smoothing out temporal fluctuations due to bursty star formation on local scales and the short-lived nature of the Hα emission. This is demonstrated by the accuracy of the exponential fit viewed as 1D azimuthally averaged radial profiles.

We fit Sérsic models convolved to the *HST* PSF, as derived by the 3D-*HST* team (Skelton et al. 2014), to postage stamps of each galaxy in F160W and F125W bands. Fitting F160W data, the centroid, ellipticity, and position angle are left free, in addition to the effective (=half-light) semimajor axis radius r_{e}, Sérsic index n_{Sersic}, and normalizing surface brightness. Initial guesses for fit parameters are taken from the fits of van der Wel et al. (2014a, hereafter vdW14) using the GALFIT software (Peng et al. 2010).[20] To estimate initial parameters for galaxies that were not fit by GALFIT (flags of $\geqslant$2 in the vdW14 catalog), we use the SExtractor parameters from Skelton et al. (2014) and empirical relations between SExtractor and GALFIT parameters for size, axis ratio ($q = 1 - \epsilon$, where ϵ is the ellipticity), and position angle. This includes empirical fits for size and axis ratio to those objects that were fit well by GALFIT, using the SExtractor major- and minor-axis size parameters A_IMAGE and B_IMAGE:

$$r_{\mathrm{e}} = 1.25 \times \sqrt{\mathrm{A\_IMAGE}^2 - 0.15^2}$$
$$q = \sqrt{B\_\mathrm{IMAGE}^2 - 0.15^2} \,/\, \sqrt{A\_\mathrm{IMAGE}^2 - 0.15^2}. \quad (5)$$

To avoid biased fits due to neighboring galaxies, we simultaneously fit all neighboring galaxies within 5″ and less than 3.5 mag fainter than the primary source in both bands (F160W and F125W), or within 2″ and less than 5 mag fainter. The left panel of Figure 3 demonstrates the good agreement of effective radii fit using IMFIT and GALFIT. The few outliers mainly move along the degeneracy between n_{Sersic} and r_{e} and tend to include multiple simultaneously fit objects of similar magnitude.

We also fit the KMOS continuum image and bootstrap realizations (a resistant weighted mean along the wavelength axis as described in Section 3.2) with the same set of constraints, resulting in similar fits, though the lower S/N leads to a larger scatter about the 1:1 relation (right panel of Figure 3).

Finally, we fit the Hα flux image (precisely, $f_{\mathrm{H}\alpha,\mathrm{WINcor}}(x,y)$) and bootstrap realizations. The Hα disk is modeled with a simple exponential profile convolved with the KMOS PSF and with the KMOS pixel size and FOV. We only fit data from spaxels with at least 20% of the nominal number of exposures for each object.[21] Unlike continuum fits, we only fit the primary galaxy (as in all but a few cases the redshift of any photometric neighbors puts any emission line outside our windows for Hα or continuum). The centroid, ellipticity, and position angle are fixed to those measured at *HST* native resolution in the F160W band to ensure that we are geometrically tracing the same disk in star formation as we see in stars. The half-light radius starting guess is fixed to that of the stars: we confirm that this has no influence on our results by refitting each source (excluding the bootstraps) with the slower differential evolution solver

[19] http://www.mpe.mpg.de/~erwin/code/imfit/

[20] We initialize the centroids to the best-fit ones from GALFIT, not SExtractor as published (A. van der Wel, private communication).

[21] For four objects in our sample, this threshold would result in spaxels with less than five individual exposures; in this case we used the latter value to define the spaxels to be fit.

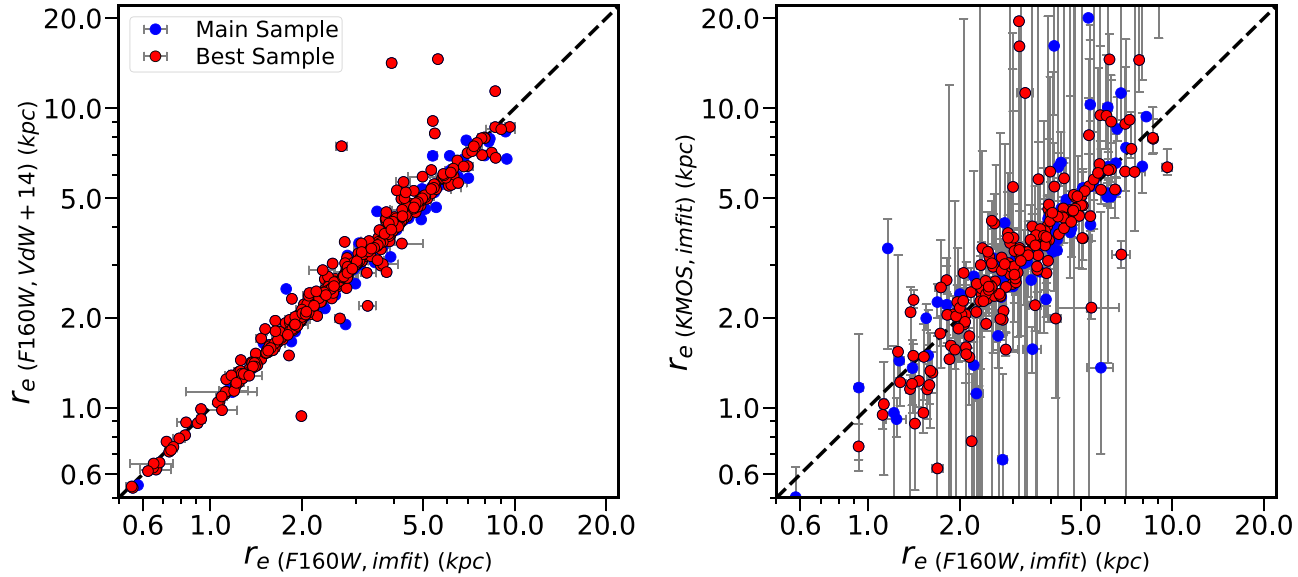

Figure 3. Measurement of intrinsic half-light size for valid fits (see text). Left: IMFIT-based fits to F160W-band CANDELS data at native *HST* resolution (this paper) compared to GALFIT fits from vdW14. Medians are within 0.2% for main and best samples, and 68% and 95% of galaxies within $^{+3}_{-5}$% and ±21%, respectively. Right: IMFIT fits at full *HST* resolution vs. those from the collapsed KMOS continuum (with bootstrap errors; see Appendix A). Fits are remarkably consistent: medians are within 1%, 68% of galaxies are within ±23%, and outlying points tend to have large error bars.

method, which does not require initial estimates for the parameters. For all galaxies in our analysis (see Section 5.1) the resultant sizes are identical. Figure 4 shows an example of this fitting procedure, and the resulting major axis Hα profile, for a galaxy in our sample. A minority of galaxies do not host such star-forming disks but do still host Hα emission tracing some other component (e.g., outflows), which not only is physically disassociated from the disk but also does not share its inclination and geometry. Cases that are not well modeled by the exponential disk model are flagged as described in Section 5.1. We further investigate the impact of our assumption of an exponential profile for the Hα emission by fitting Sérsic models to these images. We find that, for most of the sources in our analysis, the best-fit Sérsic index is close to unity and the effective radii are consistent with those from the exponential fits. Quantitatively, we find consistency between r_e from Sérsic and exponential fits within 1σ, 2σ, and 3σ for 65%, 75%, and 86% of the galaxies in our sample, respectively. We also find that the trends presented in Section 6 are unchanged within the uncertainties. However, a substantial number of Hα images have Sérsic best fits that hit the fitting limits for the Sérsic index, leading to more discrepant values for r_e when compared to the exponential fits. For these reasons, in this work we use the results of the exponential fits to the Hα images.

Appendix A describes how we test and derive final errors on size measurements, concluding not only that the continuum sizes are consistent with those from *HST*, as shown in Figure 3, but also that deviations are, statistically, very consistent with our errors, and that the accuracy of our PSF model is not the dominant source of error.

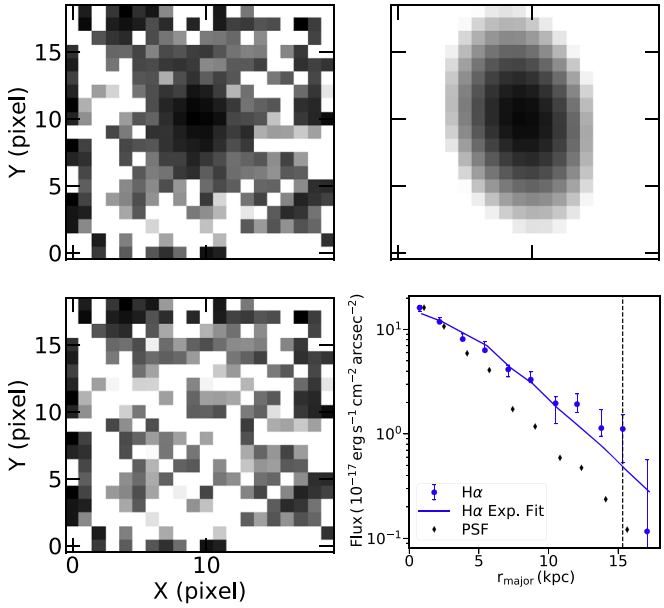

Figure 4. Example IMFIT fit to the data for galaxy GS3_11606. Top left: log-scaled Hα image; top right: best-fit model exponential galaxy convolved with PSF, from IMFIT (log-scaled); bottom left: residual image (linear scaling); bottom right: log-scaled radial profile, extracted using elliptical apertures from both data (blue points with 1σ bootstrap errors) and best-fit model (blue solid line) images. For comparison, the PSF image is also extracted in the same apertures (black diamonds). The residual image and 1D profile demonstrate a good fit for this galaxy, with a best-fit size of $r_e = 3.74^{+0.6}_{-0.45}$ kpc. The vertical dashed line indicates the radius where the major axis first crosses the edge of the KMOS FOV.

4.3. Major-axis Radial Profiles

For our current purposes we are not interested in azimuthal variations. Therefore, to establish that the radial flux profiles are indeed well characterized by the model and fit, we also extract one-dimensional radial surface brightness profiles in elliptical annuli, aligned to the galaxy's best-fit ellipticity and position angle. In practice, these are computed by embedding the images in a larger grid, weighting the individual image pixels by their effective contribution per radial bin, and then integrating the flux within differential elliptical bins corresponding to a given major-axis-equivalent radius. Equivalent profiles generated from each bootstrap image provide estimates

of the error on the profile. We also generate profiles for the best IMFIT fit model profiles in the same way—the projection of the model image onto elliptical radial apertures, and of the PSF (placed at the centroid of the galaxy). This enforces the same pixelization as exists in the data themselves for direct comparison. In Appendix B we show a gallery of Hα profiles spanning a range of size, redshift, and observed surface brightness, to highlight the quality of the data and of the fitting procedure.

5. Sample

5.1. Flagging

Our analysis requires size estimates that accurately reflect the true profiles of continuum and Hα. Starting with our astrometrically calibrated, Hα-detected, and masked sample with valid velocity grown maps (457 galaxies), we need to weed out objects with strong skyline contamination or poorly fit continuum/ Hα profiles. We do this via a series of steps during which two authors (M.F. and D.J.W.) independently visually inspected the data, removing objects not satisfying a series of requirements. This results in a sample including all the objects for the analysis of the Hα and F160W sizes (the Hα sample hereafter), which we further split into a MAIN and a BEST sample, where the BEST sample includes only the best IMFIT fits and sky subtraction (for Hα), while the MAIN samples include all reasonable fits.

In detail, we first inspected the data for atmospheric skyline residual contamination, which is evaluated by simultaneously inspecting the inverse-variance weighted summed spectrum within the mask of good Gaussian fits, the variance spectrum, and the inverse-variance weighted summed spectrum from outside the mask. The comparison of spectra inside and outside the mask serves to establish which spectral features are associated with the galaxy spectrum and which are spurious features commonly associated with high-variance residuals from sky subtraction. A total 83 of 457 galaxies have strong skyline contamination and will be excluded from further analysis, while 101 have a weaker contamination (skyline residuals are subdominant compared to the underlying Hα emission, or lie just outside the Hα wavelength range). The latter are included in the analysis as part of the Hα MAIN sample, but we examine their influence on our results by excluding them from the Hα BEST sample.

We also exclude 22 of the remaining galaxies owing to close pairs for which the paired galaxy also appears in the Hα image, 2 galaxies with PSF axis ratios >1.5, and 6 galaxies for which the apparent, measured Hα flux is incoherently spread across the FOV.

In general, our fits are deemed to be good if (a) the magnitude of fractional residuals and χ^2 in the image plane is visually defined to be small, (b) the 1D extracted profiles from data and best fit are visually in close agreement (and generally within the error bars in radial bins),[22] and (c) IMFIT converged on a best fit that did not hit the parameter limits. For Sérsic fits to the continuum, the parameters are limited to the ranges $0.2 < n_{\rm Sersic} < 8.0$, surface brightness $I(r_e) > 0$, 0.01 kpc $< r_e <$ 99 kpc, and, for exponential fits to Hα, central surface brightness $0 < I_0 < 10.I_{\rm peak}$, where $I_{\rm peak}$ is the maximum surface brightness per pixel, and 0.01 kpc $< r_e <$ 99 kpc. A total of 399 of 457 galaxies meet these criteria, of which 207 fits are excellent.

At native CANDELS resolution in F160W, 559 of 645 galaxies are well fit with a single Sérsic profile. We also require best-fit CANDELS F160W ellipticities $\epsilon < 0.7$ (which corresponds to an inclination of $i \gtrsim 72.5$ in an infinitely thin disk) because for the edge-on cases the intrinsic disk thickness cannot be ignored.

Merging all these criteria, our final Hα MAIN sample contains 281 galaxies, with 89 in the Hα BEST sample. For the remainder of the paper we will discuss the results derived for the Hα MAIN sample only, having tested that there are no qualitative changes if we restrict to the Hα BEST sample.

Of the final Hα MAIN (BEST) galaxy sample, 42 (11) have clear broad lines, and 38 (13) have known AGNs. We do not remove these from the samples, as these galaxies have been screened for a good fit exponential profile, suggesting that the outflow/AGN does not dominate the profile. Results are examined with and without these objects, with no notable difference to our conclusions.

The samples for comparison of CANDELS F160W and KMOS continuum sizes are slightly different, requiring good (excellent) fits in both continuum bands and ellipticities $\epsilon < 0.7$. This results in sample sizes of 288 and 193 for the continuum MAIN and continuum BEST samples, respectively. Galaxy sizes and errors for the Hα MAIN sample are provided in Appendix C for the online version of the article.

5.2. Sample Bias

Figure 5 examines whether any bias might be introduced through galaxies remaining undetected or unmapped in Hα or otherwise not included in the Hα MAIN or BEST samples. The top panel shows the difference between a galaxy SFR and the MS SFR of a galaxy at the same mass and redshift using the prescription derived by Whitaker et al. (2014). We show only galaxies observed by KMOS^{3D}; see Wisnioski et al. (2015), Wuyts et al. (2016), and W19 for the impact of our *Ks*-band selection in this plane. All 645 target KMOS^{3D} galaxies are shown in Figure 5. The middle panel shows the location of the same galaxies in the F160W half-light size—stellar mass plane. The bottom panel shows the galaxies in the "*UVJ*" ($U-V$ vs. $V-J$) color–color plane (Williams et al. 2009), which separates passive galaxies from star-forming galaxies (at the black demarcation line). Using two colors, it is possible to disentangle passive galaxies (top left) from optically red, dusty star-forming galaxies (top right).

Figure 5 shows that the fraction of galaxies with detected Hα is high for MS galaxies and then drops rapidly to low SFR and redder colors. Galaxies designated "unmapped" in Figure 5 can have either low S/N or chaotic kinematics/dominant broad-line components, and it is impossible to trace a dominant star-forming disk-like component. Combining nondetections and unmapped galaxies accounts for 83% of *UVJ* passive galaxies and 98% of galaxies more than 1.0 dex below the MS, but only 14% of *UVJ* star-forming galaxies and 11% of galaxies less than 0.3 dex below the MS. Of this, the unmapped population contributes an increasing fraction at high mass, reaching ∼20% for $\log_{10}(M_*/M_\odot) > 10.9$ galaxies within 1 dex of the MS. These very massive objects are more difficult to map in Hα owing to their lower specific SFRs and higher dust extinctions, which in turn make their redshift determinations more uncertain. However, the number of these objects is relatively low, and we found no

[22] Note that not all radial bins have to be consistent with the fit: our flag represents an *average* and is meant to indicate whether the size measurement is likely to be biased.

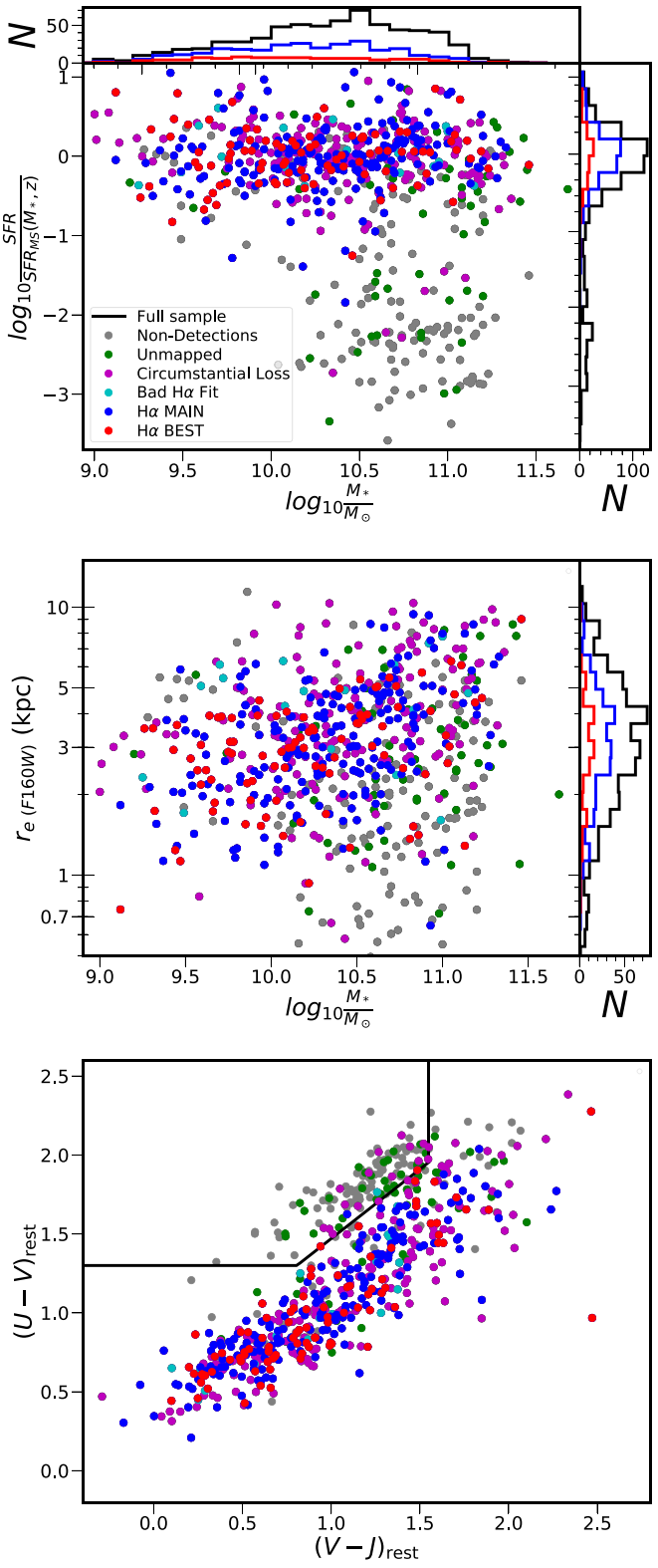

Figure 5. Top panel: offset in SFR of each KMOS^{3D} galaxy from the MS at its stellar mass and redshift, as defined by Whitaker et al. (2014); middle panel: continuum (F160W) size vs. stellar mass; bottom panel: $U-V$ vs. $V-J$ color–color space (passive galaxies are to the top right of the black demarcation line). The histograms show the marginalized distributions for the full, Hα BEST, and Hα MAIN samples in stellar mass, SFR, and galaxy size. Galaxies in the Hα BEST and Hα MAIN samples occupy the whole MS, falling off to low SFR and into the passive region of the color–color space. These samples are therefore representative of normal MS galaxies, with no significant detection bias.

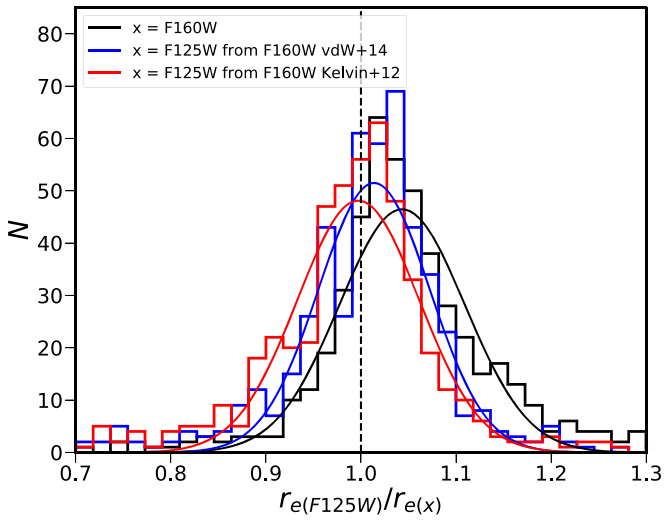

Figure 6. Histograms of the ratio of the best-fit half-light size in the F125W band to that in the F160W band for galaxies in the CONTINUUM MAIN sample, and a Gaussian best fit (black lines). The blue and red lines show the ratio with observed F160W sizes corrected to F125W with the fitting functions of vdW14 and Kelvin et al. (2012), respectively. The black dashed line marks where the size ratio is unity.

significant difference in the distributions of size and offset from the MS for the Hα MAIN sample if we split above and below $\log_{10}(M_*/M_\odot) = 10.5$. The Hα MAIN sample probes well into the dusty star-forming region of the *UVJ* diagram, which is a key feature of the KMOS^{3D} survey design (see Wisnioski et al. 2015, 2019) that helps in reducing selection biases for our sample. Along the MS, most galaxies that are not in the Hα MAIN sample are dropped for circumstantial reasons, which does not introduce any selection bias (magenta points), i.e., mostly due to significant skyline residuals or high ellipticity. The cyan points denote the few galaxies with clean, mappable emission-line signal for which the exponential fit to Hα emission was flagged as bad. There are 18 such objects, of which the majority display Hα emission offset from the continuum and some are clear mergers. These galaxies tend to have large continuum sizes for their stellar mass (bottom panel), but given their small number and, for some of them, uncertain continuum sizes, we consider there to be no notable bias against normal disk-like extended star-forming disks in our sample. Finally, we note that there is no notable difference in the properties of Hα MAIN and Hα BEST galaxies supporting our decision to focus on the Hα MAIN sample.

6. Results

6.1. Hα Size Correlations with Continuum Size and Stellar Mass

Armed with accurate galaxy half-light (size) measurements in continuum (tracing stars) and in Hα (tracing star formation), we now examine how the extent of star-forming gas relates to other known galaxy properties. In particular, we are interested in how closely the distribution of new stars, as traced by Hα, follows the distribution of old stars. Before answering this question, we turn our attention to the best tracer for the size of old stars. Kelvin et al. (2012) and vdW14 showed that the star-forming galaxies have negative color gradients, implying that their size is smaller at longer rest-frame wavelengths. This is shown in the black histogram of Figure 6, where we plot the

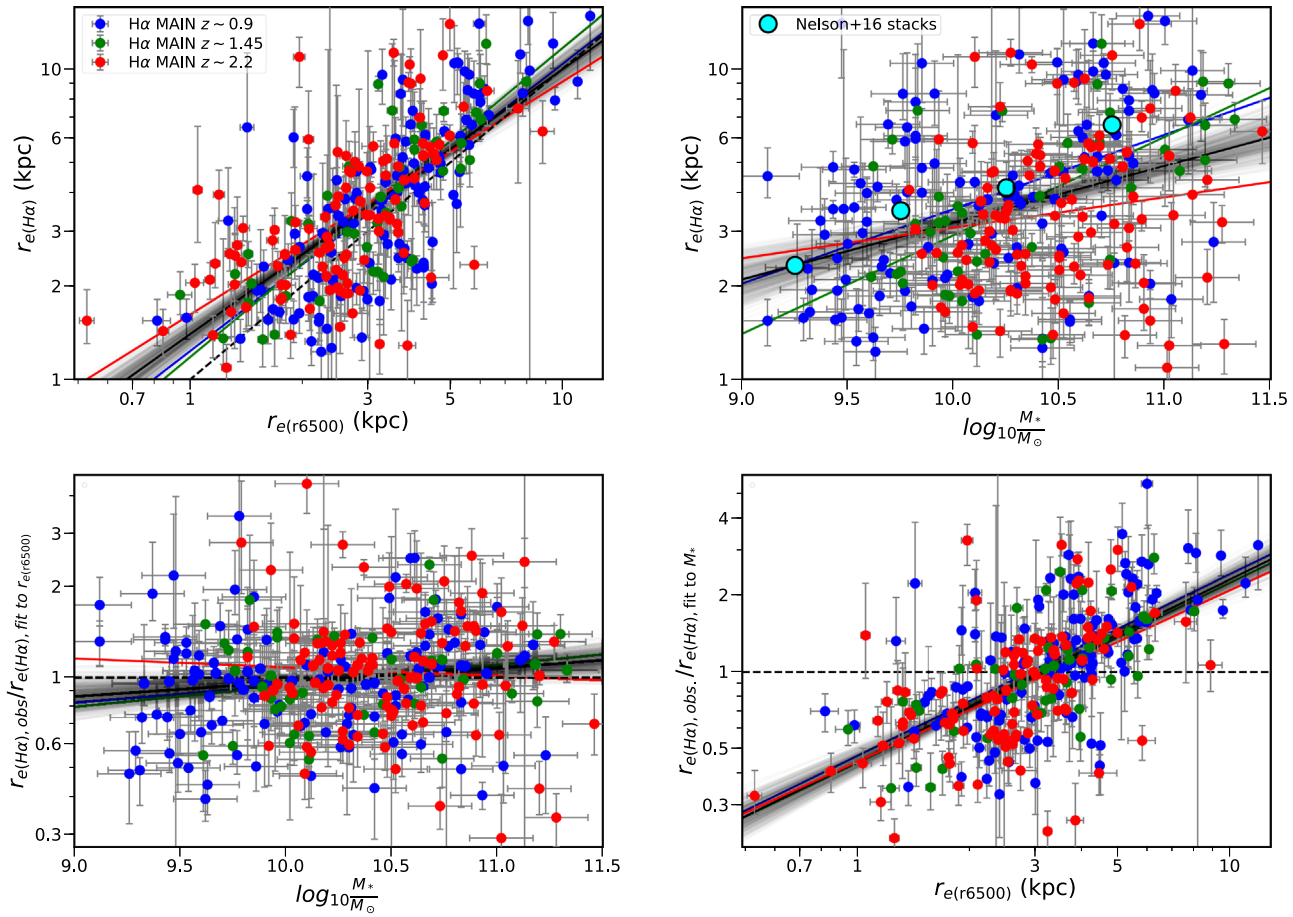

Figure 7. Top panels: Hα galaxy size plotted against galaxy size measured at 6500 Å rest-frame band (from F160W data corrected with the Kelvin et al. 2012 function; left panel) and galaxy stellar mass (right panel). The best fit and sample fits from MCMC are shown with the black solid line and fainter gray lines. Galaxies are divided into three redshift bins (corresponding to the KMOS band for Hα observation), with data points and fits to each redshift bin shown with different colors. Points from the stacked, circularized Hα size measurements of Nelson et al. (2016a), assuming an average ellipticity of $\epsilon = 0.4$, are shown in bins of stellar mass. Bottom panels: residual Hα size after subtraction of the best-fit relation above, plotted against the other parameter. The dashed black horizontal line shows the level at zero residuals. The Hα size of a galaxy is more tightly correlated with its continuum size than with its stellar mass.

ratio of sizes in F125W to F160W. The center of the Gaussian fit assumes a value of 1.042 for the size ratio. Kelvin et al. (2012) and vdW14 provided fitting functions for the wavelength dependence of the observed size, which we applied to our observed F160W sizes to correct them to F125W. We note that the Kelvin et al. (2012) correction is only a function of observed wavelength, while the vdW14 correction depends also on the galaxy stellar mass. The blue and red histograms and Gaussian fits show that these corrected sizes match the observed F125W sizes much better than the uncorrected data, with average ratios of 1.013 and 0.997 for vdW14 and Kelvin et al. (2012), respectively. Due to the simpler nature of the correction proposed by Kelvin et al. (2012) and its excellent accuracy in correcting the sizes in our sample, we use this fitting function to correct the observed F160W sizes to rest-frame 6500 Å. This rest-frame wavelength has multiple advantages: first and foremost it is close to the rest-frame wavelength probed by F160W in the center of our redshift range, and therefore a roughly equal number of galaxies are corrected to a shorter and longer wavelength. Moreover, it is close to the rest-frame wavelength of the Hα emission, mimicking the observing strategy of narrowband surveys in the local universe (where the continuum size is evaluated from a filter close in wavelength to the narrowband filter used for the Hα observations, as done by, e.g., Fossati et al. 2013; Boselli et al. 2015). We add scatter to the correction applied to individual galaxies by randomly sampling a Gaussian function, with σ equal to the standard deviation of the Gaussian fit shown in Figure 6. Hereafter we will use this corrected continuum size as a tracer for the size of the old stars, and we label it $r_{e(r6500)}$, unless otherwise noted.

At high redshift, the high gas masses, densities, and accretion rates of KMOS^{3D} galaxies mean that their star-forming gas is mostly molecular. Well-defined global galaxy relationships such as that between total star formation and stellar mass (the star-forming MS) exist primarily because total SFRs, to first order, smoothly track accretion rates, which themselves depend mainly on the global halo potential and growth (e.g., Bouché et al. 2010; Lilly et al. 2013). It remains unclear to what extent a strong evolution in mass and mass growth should be reflected in changes in galaxy size and size growth. Therefore, we begin by asking whether the Hα size of galaxies is better correlated with stellar size or total stellar mass.

Figure 7 conclusively answers this question. In the top panels, we plot the Hα size of the Hα MAIN sample against continuum size (left panel) and stellar mass (right panel). The

positive correlation between Hα and continuum size is strong and tight; a positive correlation exists also between Hα size and stellar mass, but with much larger scatter.

We fit these relations using the *linmix* package for python,[23] which follows the Bayesian framework described by Kelly (2007) and incorporates measurement errors on *x*- and *y*-axes, as well as an additional component of intrinsic scatter into the fit. This code makes use of Markov Chain Monte Carlo (MCMC) to fit a linear relation (in this case in the log–log plane and so corresponding to a power-law relation between linear quantities). A random uncertainty of 0.15 dex in $\log_{10}(M_*/M_\odot)$ is included in the analysis, consistently with vdW14.

In the size–size plane, we find a mildly sublinear best-fit relation with a slope of 0.85 ± 0.05 and intrinsic scatter of $43\% \pm 3\%$ (0.15 dex) at fixed continuum size, significantly smaller than that of continuum size versus stellar mass (56%, or 0.19 dex; vdW14). This is shown by the black solid line in Figure 7, while the gray shaded area shows a wider range of draws from the posterior. The correlation with mass has a slope of 0.18 ± 0.03 and intrinsic scatter of $68\% \pm 4\%$ and is roughly consistent with the stacked Hα size measurements of Nelson et al. (2016a), who provide fits to stacked and circularized Hα profiles from the 3D-*HST* survey in stellar mass bins (outlined cyan circles, scaled to major-axis sizes assuming an average ellipticity $\epsilon = 0.4$).

We now examine the importance of the second parameter in the bottom panels. In each case we plot the residual of Hα size with respect to the best-fit relation from the top panel against the second parameter (i.e., left, vs. stellar mass; right, vs. continuum size) and then fit relations for these residuals. Here it is clear to see that the stellar mass adds nothing to the prediction of Hα size once the continuum size has been taken into account: the residual relation fits a slope consistent with zero, and the intrinsic scatter drops only by 1% to $42\% \pm 3\%$. On the other hand, the continuum size correlates well to the residual of Hα size at fixed stellar mass, with a slope of 0.71 ± 0.05 and an intrinsic scatter of $46\% \pm 3\%$, down from $68\% \pm 4\%$ when fitting versus stellar mass only.

The stellar mass only has any relevance because it is correlated with the continuum size: once the correlation with continuum size is removed, then there is no residual relation of Hα size with stellar mass. In other words, star formation, on average, spatially tracks existing stars, but at fixed continuum size the global amount of stars has no relevance. The scatter of Hα size with stellar mass has a larger contribution from the 55% scatter between continuum size and stellar mass[24] than from the 43% scatter between Hα size and continuum size.

We divide the Hα MAIN galaxies into three subsamples of redshift, according to the KMOS band in which we observe the Hα emission line. These are $0.58 \leqslant z < 1.04$ (blue points, *YJ* band), $1.27 \leqslant z < 1.62$ (green points, *H* band), and $1.98 \leqslant z < 2.68$ (red points, *K* band). We also derive best-fit relations for each of these subsamples (colored lines). Best-fit slopes to the continuum size– Hα size relation of 0.93 ± 0.08 (*YJ*), 1.00 ± 0.11 (*H*), and 0.75 ± 0.09 (*K*) are consistent with our combined best-fit relation within 2σ, and with close to a linear relation (with the possible exception of the highest redshift bin, which is 2.8σ away).

The fit intercept corresponds to the typical Hα size of a galaxy at a particular continuum size: for a near-linear relation this is a near-constant ratio. We derive this ratio at the median continuum size of 3.23 kpc to be 1.18 ± 0.03, i.e., a median Hα size that is 18% larger than the continuum size. Folding in the measured intrinsic scatter, this corresponds to a mean Hα size that is 26% larger than the continuum size.[25] This compares to the typical size ratio of ∼1.3 found by Nelson et al. (2012) in highly star-forming 3D-*HST* galaxies at $z \sim 1$ and the median ratio from stacked profiles of normal star-forming 3D-*HST* galaxies (∼1.1; Nelson et al. 2016a). Within our wide redshift range, there is no evidence for evolution in this value, with consistent best-fit values of 1.13 ± 0.05, 1.17 ± 0.06, and 1.20 ± 0.05 in the three redshift bins defined above. So while Hα sizes do track existing stellar sizes, they are, in a median (mean) sense, larger by ∼18% (∼26%) over our full redshift range. That star formation sizes are larger than stellar sizes is a precondition for in situ size growth, and we shall return to this topic in Section 7.2.

6.2. Which Other Parameters Influence Hα Size?

Figure 7 demonstrates that, at fixed galaxy continuum size, there is no significant residual dependence on redshift (when split into three bins). This is confirmed by directly fitting the residual to the Hα size—continuum size best fit against redshift (Figure 8, top left panel). We find only a marginal dependence on redshift at $<2\sigma$ level, and the intrinsic scatter of the size–size relation does not drop when redshift is included as a third parameter.

In Table 1 we show the best-fit slope β,[26] the fraction of Monte Carlo realizations with β greater than zero ($P(\beta > 0)$), the Spearman rank correlation coefficient (ρ), and the probability that it is consistent with the null (no-correlation) hypothesis ($P(\rho|\text{null})$) for the residuals of the size–size relation versus several other parameters.

Having explored the role of continuum size, stellar mass, and redshift, we now turn to examining the dependence on SFR and other quantities that are known to correlate with it. This is done in Figure 8, where we plot the size–size relation residuals versus the SFR in the top right panel and specific SFR in the middle left panel. In the middle right panel, we look at the residuals versus the ratio of inferred molecular gas mass to stellar mass, $M_{\rm molgas}/M_*$, estimated using the relation of Tacconi et al. (2018, their BEST sample as in their Table 3 (b)), which depends on z, δ(MS) (logarithmic offset from the Whitaker et al. 2014 MS relation), and M_*. There is no significant trend in residual size with SFR, sSFR, δ(MS), or inferred $M_{\rm molgas}/M_*$, nor any notable decrease in scatter, suggesting that—at least for normally star-forming MS galaxies—the star-forming gas traces the stars in exactly the same way independent of the relative amount of star-forming gas or of star formation efficiency within the limits to which we can measure it.

In the bottom panels of Figure 8 we examine the dependence on galaxy morphology parameterized by the Sérsic index of our fit in the F160W band and the bulge-to-total ratio B/T from the fits of Lang et al. (2014). The sample of Lang et al. (2014) partially overlaps with the Hα MAIN sample, with only 60%

[23] https://github.com/jmeyers314/linmix

[24] This value derived from our sample is fully consistent with the 56% found by vdW14.

[25] The mean of a lognormal distribution = median+$0.5\sigma^2$.

[26] With the symbol β we simply refer to the slope of the power-law fit, and this parameter has no relation with the UV β slope.

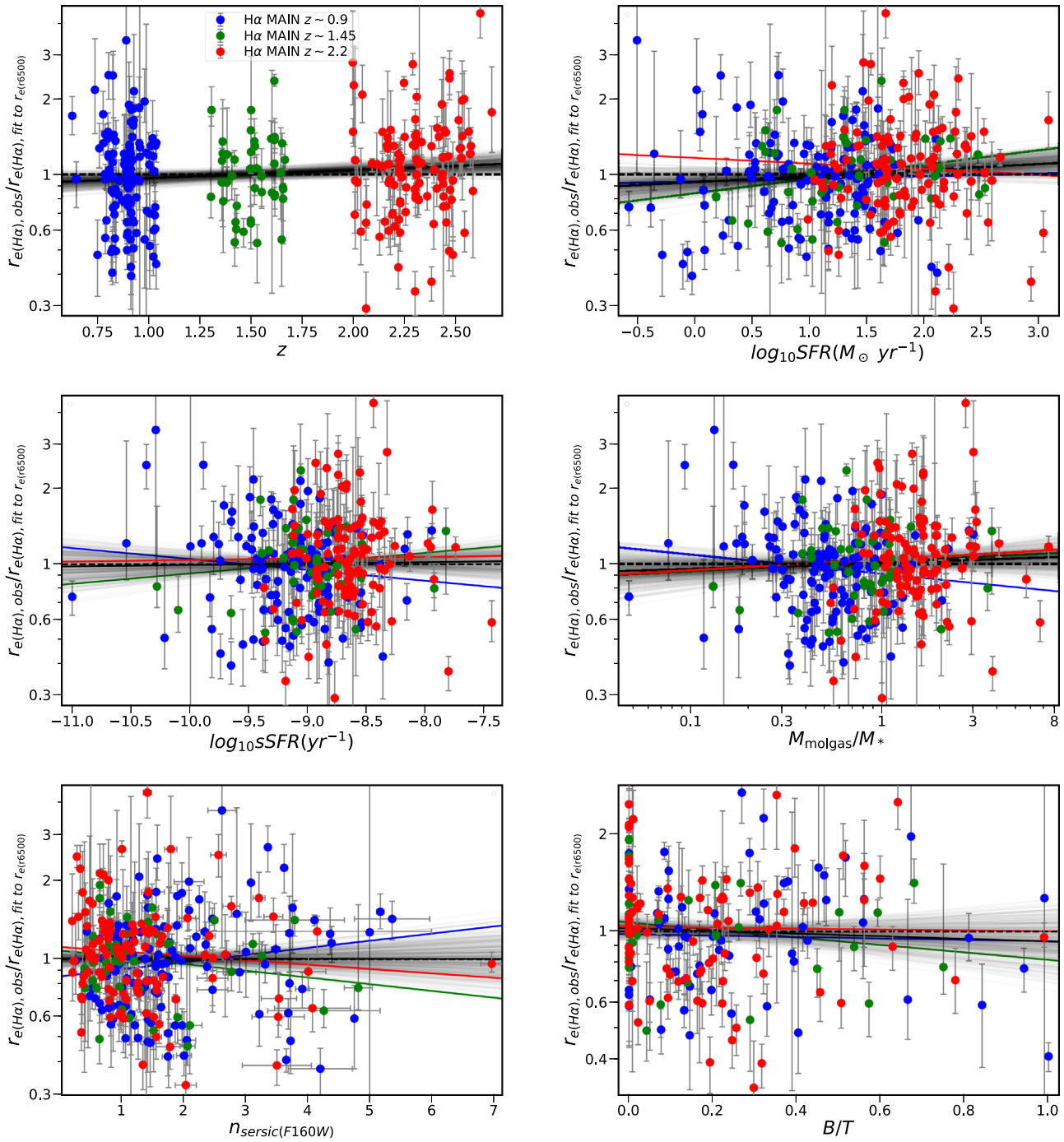

Figure 8. Residual Hα size with respect to the best fit vs. continuum size (see Figure 7 top left panel) vs. redshift (top left), SFR (top right), specific SFR (middle left), ratio of molecular gas to stellar mass (as inferred from Tacconi et al. 2018; middle right), Sérsic parameter measured in the F160W band (bottom left), and bulge-to-total ratio (as measured by Lang et al. 2014 for a subset of our Hα MAIN sample; bottom right). The best fit and sample fits from MCMC are shown with the black solid line and fainter gray lines, respectively (and colored lines for the three independent redshift bins). The black dashed line shows the level at zero residuals.

of the galaxies having a valid value of B/T. Nonetheless, this subsample is large enough to derive statistically robust conclusions. As demonstrated in the figure and in the table, there is no significant dependence. This contrasts with the situation in the local universe, where normally star-forming (gas-rich) galaxies with little or no bulge tend to have very similar Hα and continuum sizes, whereas those with more significant bulges have relatively larger Hα sizes (Fossati et al. 2013). In galaxies with more bulge, the half-light size of the combined bulge+disk is less than that of the disk, and the Hα-emitting gas usually shows little sign of a bulge component. Therefore, it is perhaps surprising that our KMOS^{3D} galaxies with higher Sérsic index or higher B/T measurements show no indication of relatively larger Hα sizes. Morphology does not seem to play a role in driving the relative size of the Hα disk in the high-redshift universe, at least not measurably within our star-forming sample.

In Table 1 we test even more parameters, including those related to the dust extinction, the galaxy color, the ratio of SFR to Hα luminosity, the ratio [N II]/Hα, and the galaxy

Table 1
Correlation of Offset from $r_e(\mathrm{H}\alpha)$–$r_e(r6500)$ Relation (Figure 7, top left panel) with Other Parameters

PARAMETER, X	β	$P(\beta > 0)$	ρ	$P(\rho\|\mathrm{null})$
z	0.03 ± 0.02	0.978	0.115	0.05
$\log_{10}(M_*/M_\odot)$	0.04 ± 0.02	0.974	0.114	0.05
$\log_{10}(\mathrm{SFR}(M_\odot\ \mathrm{yr}^{-1}))$	0.02 ± 0.02	0.921	0.088	0.14
$\log_{10}(\mathrm{sSFR}(\mathrm{yr}^{-1}))$	0.01 ± 0.02	0.591	0.037	0.53
$\log_{10}(\delta(\mathrm{MS}))$[a]	-0.01 ± 0.03	0.424	−0.015	0.80
$\log_{10}(M_{\mathrm{molgas}}/M_*)$[b]	0.02 ± 0.03	0.767	0.075	0.21
$n_{\mathrm{Sersic,F160W}}$	0.00 ± 0.01	0.496	−0.005	0.93
B/T, F160W[c]	-0.04 ± 0.06	0.249	−0.020	0.79
A_V	-0.03 ± 0.02	0.059	−0.08	0.18
$\log_{10}(L_{\mathrm{H}\alpha,\mathrm{erg\ s}^{-1}})$ from fit[d]	0.11 ± 0.02	1.0	0.31	$<10^{-5}$
$\log_{10}(\frac{\mathrm{SFR}(M_\odot\ \mathrm{yr}^{-1})}{L_{\mathrm{H}\alpha,\mathrm{erg\ s}^{-1}}})$ from fit[d]	-0.10 ± 0.02	0.0	−0.22	0.0001
$\log_{10}(\frac{\mathrm{SFR}(M_\odot\ \mathrm{yr}^{-1})}{L_{\mathrm{H}\alpha,\mathrm{erg\ s}^{-1}}})$ from fit[d] and dust-corrected[e]	-0.11 ± 0.03	0.0	−0.23	0.0001
$(U-V)_{\mathrm{rest}}$	-0.09 ± 0.06	0.080	−0.072	0.25
$(V-J)_{\mathrm{rest}}$	-0.09 ± 0.05	0.031	−0.091	0.13
ϵ[f]	-0.13 ± 0.14	0.171	−0.122	0.04
$\log_{10}([\mathrm{N\ II}]/\mathrm{H}\alpha)$[g]	0.02 ± 0.07	0.594	0.074	0.34
$\log_{10}(\delta_{0.75})$[h]	-0.01 ± 0.03	0.36	−0.07	0.37
$\log_{10}(M_{h\|\mathrm{cen},50\%}/M_\odot)$[i]	0.05 ± 0.03	0.96	0.09	0.12
P_{sat}[j]	-0.08 ± 0.05	0.07	−0.07	0.23

Notes. Best-fit between the parameter given in the first column (X) and the residual Hα size relative to the best fit versus continuum size for the full Hα MAIN sample $\left(Y = \frac{r_{e(\mathrm{H}\alpha),\mathrm{obs}}}{r_{e(\mathrm{H}\alpha),\mathrm{fit}}}\right)$. Fits are of the form $Y = 10^{\alpha+\beta.X}$, with the slope β and 1σ errors given in Column (2). In Column (3), $P(\beta > 0)$ is estimated from the fraction of MCMC realizations with a positive slope, so values around 0.5 are random, while values of 0 or 1 indicate a significant negative or positive slope, respectively. Column (4) contains the Spearman rank correlation coefficient, and Column (5) indicates the probability (p-value) of such a value assuming no correlation between the two parameters.

[a] $\delta(\mathrm{MS}) = \mathrm{SFR}(M_\odot\ \mathrm{yr}^{-1})/\mathrm{SFR}_{\mathrm{MS},z,M_*}(M_\odot\ \mathrm{yr}^{-1})$, with the parameterizations of the MS from Whitaker et al. (2014).

[b] Inferred via $\log_{10}(M_{\mathrm{molgas}}/M_*) = -1.25 + 2.6\log_{10}(1+z) + 0.53\log_{10}(\delta(\mathrm{MS})) - 0.36\log_{10}(M_*)$ (Tacconi et al. 2018, best-fit relation).

[c] Restricted to galaxies with valid B/T measurements in the F160W band from Lang et al. (2014).

[d] $L_{\mathrm{H}\alpha,\mathrm{erg\ s}^{-1}} = 2\pi.h^2.I0.(1.-\epsilon)$.

[e] Using the differential dust recipe of Wuyts et al. (2013): $\log_{10}(L_{\mathrm{H}\alpha,\mathrm{erg\ s}^{-1},\mathrm{dust.cor}}) = \log_{10}(L_{\mathrm{H}\alpha,\mathrm{erg\ s}^{-1}}) + 0.4(1.9A_{\mathrm{cont}} - 0.15A_{\mathrm{cont}}^2)$.

[f] Galaxy ellipticity, ϵ from fit to F160W band.

[g] Restricted to galaxies in the Hα MAIN sample for which the skyline residual contamination at the wavelength of the red [N II] emission line is not strong.

[h] Environmental overdensity in 0.75 Mpc apertures from Fossati et al. (2017).

[i] 50th percentile of probability distribution function for halo mass assuming that the galaxy is the central of its halo from Fossati et al. (2017).

[j] Probability galaxy is a satellite, as calibrated by Fossati et al. (2017).

environment (parameterized by the local galaxy overdensity, halo mass, and probability of being a satellite galaxy from Fossati et al. 2017).

Our analysis shows that most parameters show little or no correlation with the residual Hα size, and the only significant trends are those related to dust extinction, which we discuss further in Section 6.4. None of these results are significantly changed if we restrict ourselves to the Hα BEST sample, eliminate galaxies hosting AGNs or broad-line emission, or

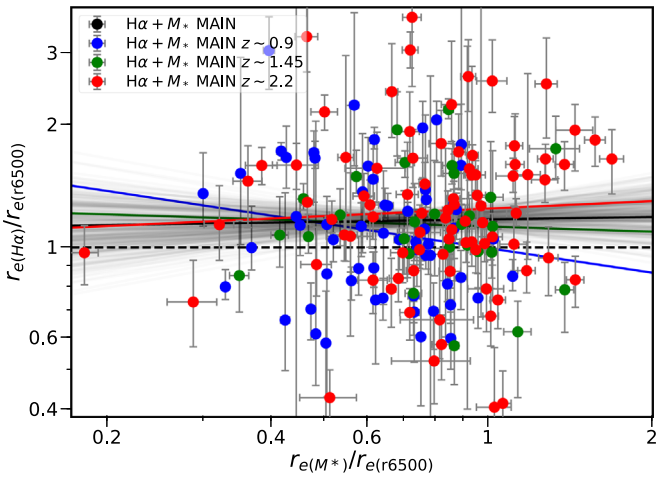

Figure 9. Ratio of the half-mass (from Lang et al. 2014) to the half-light size vs. the ratio of the Hα size to the half-light size. Points and lines are color-coded as in Figure 8. The lack of a significant correlation demonstrates that whatever drives variations in $r_{e(M_*)}/r_{e(r6500)}$ is not driving variations in $r_{e(\mathrm{H}\alpha)}/r_{e(r6500)}$.

limit to the stellar mass range of our highest redshift bin ($\log_{10}(M_*/M_\odot) \gtrsim 10$).

6.3. Caveats

In detail, the picture is more complicated. While much of the Hα emission locally traces the ionizing star formation, there can be a more diffuse component, additional sources of ionization, and much of the emission can be obscured by dust, especially in dense and infrared-bright starbursting regions. For example, two very high mass star-forming galaxies in our sample (U4_20704 and U4_36247 at $M_* > 10^{11}\,M_\odot$) have been observed by ALMA at 870 μm (Tadaki et al. 2017) to have significantly smaller submillimeter sizes than their Hα or F160W sizes.

Also the continuum emission, which we have corrected to rest-frame 6500 Å to remove the effects of a variable rest-frame wavelength for galaxies at different redshifts, does not perfectly trace the stellar mass in the galaxy. We examine the Sérsic profile fits to stellar mass maps produced by Lang et al. (2014) for $0.5 < z < 2.5$ and $M_* > 10^{10}\,M_\odot$ galaxies in our sample. We find for $n_{\mathrm{Sersic}} < 2$ galaxies that half-mass sizes are on average ∼75%–80% of the rest-frame 6500 Å half-light sizes with a <5% dependence on redshift, consistent with Wuyts et al. (2012, their Figure 11). The ratio $r_{e(M_*)}/r_{e(r6500)}$ decreases for high-n_{Sersic} sources in our lower redshift bin, possibly demonstrating the influence of a higher mass-to-light ratio bulge in such galaxies. However, as shown in Figure 9, we find no correlation between $r_{e(M_*)}/r_{e(r6500)}$ and $r_{e(\mathrm{H}\alpha)}/r_{e(r6500)}$, demonstrating that the driver of variations in $r_{e(M_*)}/r_{e(r6500)}$, such as large bulges in some galaxies at lower z, is not driving variations in $r_{e(\mathrm{H}\alpha)}/r_{e(r6500)}$. This suggests that, in our observed continuum band and for the normally star-forming galaxies in our sample, we are predominantly tracing the disk and are measuring disk sizes, even in galaxies with significant bulges.

The color gradients in late-type galaxies appear likely to result predominantly from gradients in dust extinction (Pastrav et al. 2013). In this case, any dust gradient will effect the extinction gradient for Hα emission and the continuum. Under the foreground screen approximation, these are equally affected, and the ratio of half-light sizes should be unaffected.

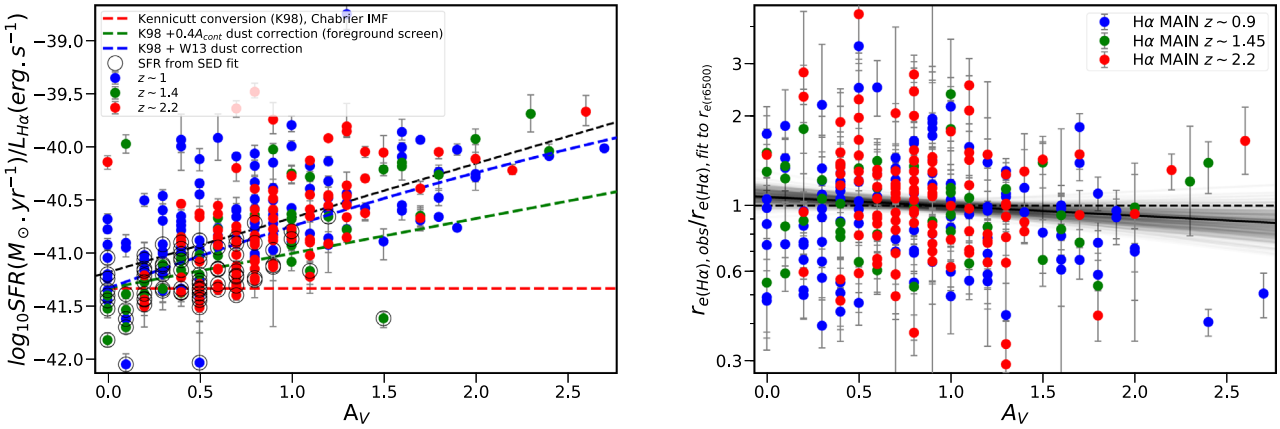

Figure 10. Left: comparison of the integrated dust obscuration of the stars with that of Hα emission tracing SFR. The *x*-axis value A_V is estimated assuming a foreground dust screen from SED fitting (Wuyts et al. 2011), while the *y*-axis is the ratio of the total estimated SFR (from IR+UV where available or SED otherwise) to the total Hα luminosity from the exponential fits. The horizontal dashed red line denotes the canonical calibration of SFR to $L_{\rm H\alpha}$ from Kennicutt (1998), assuming a Chabrier IMF of stars. Assuming that star formation is obscured to the same degree as older stars, we get the dashed green line. An additional component of extinction as empirically calibrated by Wuyts et al. (2013) is shown as the blue dashed line. This comes very close to our best-fit relation (black dashed line). Galaxies with no IR detection and therefore with SFR from SED fits are outlined with black circles: clearly these are the less dusty objects. Right: the dependence of the residual Hα size at fixed continuum size on A_V is weak and not highly significant.

However, as we shall see in Section 6.4, there is excess extinction of Hα emission associated with dust embedded in the star-forming H II regions, well described on average by differential extinction laws such as Wuyts et al. (2013). The dustier part of the galaxy (typically the center; Wuyts et al. 2012; Nelson et al. 2016b) will have extra extinction in Hα compared to continuum, implying that while both observed sizes will be larger than the true size, the Hα size should be increased by a larger amount. In other words, the ratio of Hα to continuum size should be an upper limit for the ratio of the size of the star-forming disk to that of the stellar mass disk.

6.4. Dependence on Dust

The left panel of Figure 10 examines the dust correction. On the *x*-axis we plot the best-fit A_V extinction estimated as part of the SED fits to multiwavelength photometry (Wuyts et al. 2011). On the *y*-axis we plot the ratio of the total SFR to the total Hα luminosity from KMOS. The SFR is estimated by Wuyts et al. (2011) and includes obscured star formation as seen via infrared emission from *Herschel*/PACS or *Spitzer*/MIPS where detected, and it is based on SED fits where there is no infrared detection. The total Hα luminosity is estimated by integrating the IMFIT exponential disk fit out to infinity. Similar results are obtained by integrating the Hα image (these are merely noisier and truncated at the edge of the KMOS FOV).[27] The *y*-axis value can therefore be interpreted as a conversion factor from Hα luminosity to total SFR including the dust correction. A standard conversion *not including dust* from Kennicutt (1998), shifted to a Chabrier initial mass function (IMF), describes the lower envelope of the data well (horizontal dashed red line). As A_V increases, so does the dust correction to this conversion factor (effectively the mass of stars formed per number of detected Hα photons). If the Hα extinction was equivalent to the continuum extinction at the wavelength of Hα, $A_{\rm cont} = 0.82A_V$ from Calzetti et al. (2000) as in Wuyts et al. (2013) (as appropriate in the case of a foreground dust screen), we would get the dashed green line. Our best fit to the data (black dashed line) implies a steeper dependence on A_V and is in good agreement with the best-fit polynomial from Wuyts et al. (2013) (blue dashed line), implying an excess extinction $A_{\rm extra}$ for the Hα emission in H II regions such that $A_{\rm H\alpha} = A_{\rm cont} + A_{\rm extra}$ with $A_{\rm extra} = 0.9A_{\rm cont} - 0.15A_{\rm cont}^2$.

In the right panel of Figure 10 we show that there is only a weak, barely significant negative correlation between the residual Hα size at fixed continuum size and A_V, implying little or no correlation between the integrated continuum obscuration and the extent of Hα emission at fixed continuum size in the galaxy, as confirmed in Table 1. Similarly weak correlations are found with rest-frame galaxy colors. The lack of a stronger correlation contrasts with simple expectations. Galaxies exhibit a centrally peaked dust extinction profile (e.g., Wuyts et al. 2012; Nelson et al. 2016b) and an excess integrated extinction at Hα increasing with increasing A_V as observed in the left panel of Figure 10. This implies that we expect extinction effects to drive a flattening of the light profiles (larger half-light radius observed than the true one; see, e.g., Pastrav et al. 2013), and that this applies more in galaxies with higher extinction (A_V) and more in Hα than in continuum ($A_{\rm H\alpha} > A_V$). As a result, the ratio of observed Hα to continuum size should be larger than the true one, and, naively, one might expect that this effect should increase with increasing A_V. Instead, we see only a very weak—and negative—correlation of the ratio of observed Hα to continuum size (at fixed continuum size) with A_V.

In contrast, Figure 11 shows how the residual Hα size at fixed continuum size does depend on the parameters specifically related to the dust obscuration of the Hα emission. The top panel shows the variation as a function of the total obscuration of Hα emission, as parameterized via the ratio of the total SFR to Hα luminosity (the *y*-axis parameter from Figure 10). In the bottom panel we correct the Hα luminosity for the average Hα dust extinction using the well-fit prescription from Wuyts et al. (2013). Now the *x*-axis values describe the *residual dust correction* to Hα emission after applying this average correction. In both panels of Figure 11, the vertical dashed green line refers to the Kennicutt (1998)

[27] Note that these estimates are both in excellent agreement with the aperture values presented by W19, for which the clearest difference is the expected dependence on the ratio of effective radius to aperture size.

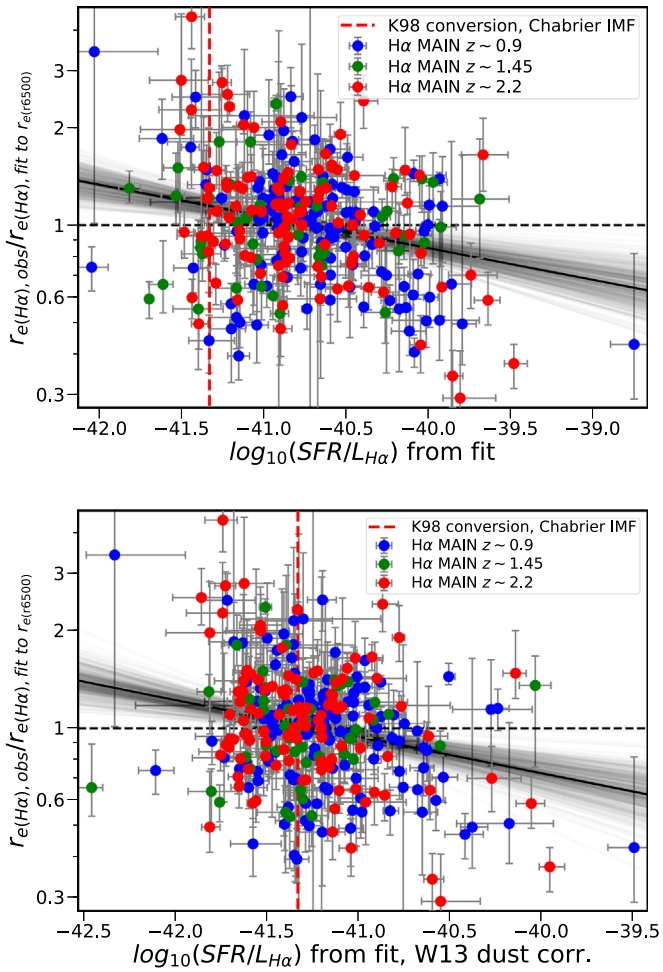

Figure 11. Residual Hα size at fixed continuum size vs. the amount of Hα dust obscuration. The top panel shows the relation to the total SFR normalized by total Hα luminosity (*y*-axis value in Figure 10, left panel). The bottom panel corrects this Hα luminosity for extinction as a function of A_V using the differential extinction law of Wuyts et al. (2013) (the blue line from Figure 10) and examines the correlation of residual Hα size at fixed continuum size with the residual Hα extinction. Even this residual dust extinction correlates significantly with the residual Hα size, indicating that galaxies with less Hα extinction at fixed A_V can be visibly more extended in Hα relative to their stars. Our best-fit relations between parameters are shown (black solid line), as well as individual Monte Carlo realizations (fainter gray lines).

conversion between SFR and (extinction-corrected) Hα luminosity.

The typical residual Hα size (at fixed continuum size) increases significantly for less total extinction of Hα, even after correction for the average global obscuration, $A_{\mathrm{H}\alpha}(A_V)$, based on the Wuyts et al. (2013) recipe. The lack of dependence of residual Hα size on continuum obscuration (Figure 10, right panel) implies that, moving from the top to the bottom panel of Figure 11, the trend is barely reduced by a global A_V-based correction. In other words, the relative Hα size at fixed continuum size must be largely independent of the total amount of foreground dust obscuration $\propto A_V$. Instead, our results must be interpreted to mean that the ratio of Hα to continuum obscuration $A_{\mathrm{H}\alpha}/A_V$ decreases for galaxies that are relatively more extended in Hα relative to stars.

We have seen no notable dependence of residual Hα size at fixed continuum size on global parameters A_V, SFR, or inferred gas fraction. That the residual Hα size *decreases* with increasing $A_{\mathrm{H}\alpha}/A_V$ must originate in the internal geometries of dust differently affecting young and old stars in galaxies. Fitting a dust model to explain correlations with the resolved extinction maps of low-redshift MaNGA galaxies, Li et al. (2019) derive a best-fitting model in which the fraction of dust in a foreground screen increases with galactocentric radius, with the rest of dust assumed to live in Hα-emitting H II regions. Such a model implies that Hα photons escape relatively more easily at large galactocentric distances than in the center of galaxies, and that galaxies with intrinsically large Hα disk sizes or steep dust obscuration profiles are less subject to extra obscuration of Hα emission.

This would put such galaxies below the average relation in the left panel of Figure 10 (blue dashed line), closer to the pure foreground screen model, and also to the left of both panels in Figure 11. These results are also tabulated in Table 1, and we find equivalent trends at similar significance if we use the total Hα luminosity from integrated Hα maps instead of the integrated exponential fit. This demonstrates that the trends are not driven by covariance in the fitting parameters.

Finally, we note that despite the details of these trends, the overall effect of dust obscuration is still to increase the observed Hα sizes more than that of the stars, with no difference only in the limiting case of no embedded dust (pure foreground screen). In other words, the measured mean (median) size ratio of 1.26 (1.18) is likely to be an upper limit on the true average ratio of the star-forming disk size to the stellar disk size. Indeed, in extreme cases from within our sample we know that there are highly star-forming massive galaxies hosting strong central star formation (Tadaki et al. 2017), which is almost completely obscured in Hα. However, it appears that the Hα sizes of such galaxies are nonetheless similar to those of their stars. Indeed, we see a relatively unobscured and subdominant component of star formation that is associated with the stellar disk, while the majority of star formation occurs in the highly optically thick central region of the galaxy, with little or no escape of ionizing photons and subsequent Hα emission.

7. Interpretation

In the previous section we have found that over the large redshift and stellar mass range probed by our observations, the ionized gas size is on average 1.26 times larger than the continuum size (tracing older stellar populations), with negligible dependence on other fundamental parameters. We now interpret these results in terms of gas accretion and its angular momentum.

7.1. Analytic Considerations

A simplified prediction for disk sizes can be derived by combining Equations (2) and (12) of Mo et al. (1998):

$$R_d \propto H(z)^{-2/3} \cdot \lambda' \cdot M_{\mathrm{halo}}^{1/3}, \quad (6)$$

for disk size R_d, galaxy spin parameter $\lambda' = \lambda \cdot \frac{j_d}{m_d}$, halo spin parameter λ, and a fraction of halo angular momentum in the disk j_d and of halo mass in the disk m_d (such that $j_d/m_d = 1$, where there is no difference in the specific angular momentum of disk and halo). R_d can then be related to the disk mass M_d via $M_{\mathrm{halo}} = (M_{\mathrm{d}}/m_{\mathrm{d}})$.

The predicted dependence on $M_{\rm halo}^{1/3}$ originates with the predicted proportionality to the circular velocity of the halo, V_c, and thus $M_{\rm halo}^{1/3}$ from the virial theorem. This is steeper than the observed dependence on stellar mass in the Tully–Fisher relation ($\sim 1/3.75 \sim 0.27$ Lelli et al. 2016), which is more in line with the slope of the stellar mass–size relation ($\sim$0.22; vdW14). A slope of less than 1/3 may be accounted for by variations in m_d with mass. The predicted dependence on $H(z)$, on the other hand, suggests a strong evolution in galaxy sizes at fixed mass. Observations are at potential odds with one another about the rate of evolution. vdW14 find that observed median sizes of star-forming disk galaxies at fixed stellar mass evolve as $H(z)^{-2/3}$, but Suess et al. (2019) claim to see little evolution at fixed stellar mass of half-mass sizes, once they account for radial gradients in the mass-to-light ratio. Both cases appear surprising, as disk sizes are likely to be set at the epoch of formation (not observation) but still evolving as they grow through star formation. That Suess et al. (2019) find a much flatter dependence on stellar mass also begs questions about the expected dependence of size on mass or circular velocity.

More straightforwardly, the measured intrinsic scatter in galaxy sizes of $1\sigma \sim 0.16 - 0.19$ dex is very consistent with the scatter in continuum galaxy size at fixed specific angular momentum (Burkert et al. 2016) and with the scatter in halo spin parameter from simulations. The latter similarity suggests that most of the scatter in galaxy sizes originates as scatter in the halo spin.

Equation (6) can also be applied to examine the star-forming disk size, noting that the parameters j_d and m_d effectively describe the efficiency of angular momentum and mass transfer from halo to disk. The term $M_{\rm halo}^{1/3}$ in Equation (6) describes the gravitational potential at the time of formation of the relevant component. Therefore, we can predict the ratio of star-forming disk size to stellar disk size:

$$\frac{R_{d,\rm SF}}{R_{d,*}} \propto \left(\frac{H(z_{\rm SF})}{H(z_*)}\right)^{-2/3} \cdot \frac{\lambda'(z_{\rm SF})}{\lambda'(z_*)} \cdot \left(\frac{M_{{\rm halo},z_{\rm SF}}}{M_{{\rm halo},z_*}}\right)^{1/3}, \quad (7)$$

where the Hubble parameter, specific angular momentum, and halo mass should be evaluated at times appropriate to the star-forming and stellar components. Naively, Equation (7) suggests that the star-forming disk should be larger than the stellar disk by an amount depending on the Hubble parameter at their relative times of formation with $z_{\rm SF} \ll z_*$, with modifications that can relate to the growth in halo mass and to changes in the halo spin parameter over time.

We found that the size in Hα emission correlates strongly with the continuum size, with an intrinsic scatter smaller than that of continuum size—and of inferred halo spin parameters—at fixed mass. This demonstrates the stability over time of the spin parameter, with less variation in time than variation between halos. The intrinsic scatter of 43% ± 3% combines any short-term temporal variation in λ with the changes in the efficiency of the angular momentum transfer from halo to disk scales, including the process of star formation, as well as the effects of dust on both size measurements. We evaluate the ratio $R = [H(z_*)/H(z_{\rm SF})]^{2/3}$ appearing in Equation (7) at redshifts $z_{\rm SF} = z_{\rm obs} = 1$ and 2, upon which our observations are concentrated. Then, with the assumed cosmology we obtain $R = 1.33$ for $z_* = 1.5$ and $z_{\rm obs} = 1$ and $R = 1.59$ for $z_* = 3$ and $z_{\rm obs} = 2$, where the z^* values are selected to have a roughly constant look-back time of 1 Gyr from the observation redshifts. On this timescale a galaxy would double its stellar mass, since $M_*/{\rm SFR} \lesssim 1$ Gyr for $z > 1$. We then evaluate the last term of Equation (7), $P = [M_{{\rm halo},z_{\rm SF}}/M_{{\rm halo},z_*}]^{1/3}$, using the halo growth factor $(1 + 1.11z) \cdot \sqrt{\Omega_M.(1+z)^3 + \Omega_\Lambda}$ given by Fakhouri et al. (2010). We get $P = 0.8(0.7)$ with the same choice of z^* and $z_{\rm obs}$ made above, respectively. In summary, Equation (7) gives $R_{d,SF}/R_{d,*} \sim 1.11(1.06) \cdot \lambda'(z_{\rm SF})/\lambda'(z_*)$, respectively. The simplest interpretation of the observed lack of evolution in the ratio of star-forming size (as traced by Hα) to that in stars (as traced in the rest-frame 6500 Å), as seen in our data, is therefore that the specific angular momentum of star-forming material is stable across many gigayears of cosmic time, resulting in an almost constant $\lambda'(z_{\rm SF})/\lambda'(z_*)$ ratio. This evidence can be physically related to the complex interplay of cooling, accretion, star formation, and feedback that might regulate the angular momentum of star-forming material and its evolution.

Table 2
The $z = 2$ Progenitor Galaxy Mass of a given $z = 1$ Galaxy in Our Toy Model

$\log_{10}\frac{M_*}{M_\odot}$ at $z = 1$	$\log_{10}\frac{M_*}{M_\odot}$ at $z = 2$	
	SF	SF + Mass Loss
10	8.02	8.63
10.5	8.76	9.38
11	9.95	10.47
11.5	10.96	11.20

Note. Stellar mass growth via star formation occurs along the MS (Whitaker et al. 2014). In the third column we include stellar mass loss in the calculation, which offsets star formation such that the mass growth is reduced. Galaxies below $M_* \sim 10^{10.8}\,M_\odot$ at $z = 1$ had stellar masses below $10^{10}\,M_\odot$ at $z = 2$, and below our approximate KMOS^{3D} limit.

7.2. A Toy Model for Evolution in Size and Mass

Observed star-forming galaxies have SFRs and stellar sizes that depend on their stellar mass and redshift, with lognormal scatter around the observed relations. In this section we construct a simple toy model to predict how star formation in galaxies should lead them to evolve in the size–mass plane.

Our model predicts the evolution in a galaxy's stellar mass by assuming it to grow purely via star formation according to the MS relation between SFR and stellar mass from Whitaker et al. (2014). We include a lognormal scatter of $\sim$0.3 dex (Noeske et al. 2007) to get estimates of mean SFR rather than median SFR, which should be appropriate assuming that individual galaxies scatter above and below the MS. This is offset by mass loss from stars computed according to the Flexible Stellar Population Synthesis code (Conroy et al. 2009). Table 2 provides the stellar mass at $z = 2$ for galaxies evolved in this way and ending with fixed masses of $\log_{10}(M_*/M_\odot) = 10, 10.5, 11, 11.5$ at $z = 1$. We also compute the much lower $z = 2$ stellar mass for a star-formation-only recipe (no mass loss, and thus much more rapid mass evolution). Galaxies below $M_* \sim 10^{10.8}\,M_\odot$ at $z = 1$ had stellar masses below our approximate KMOS^{3D} limit of $10^{10}\,M_\odot$ at $z = 2$.

Motivated by the consistent ratio of Hα to continuum sizes $r_{\rm e(H\alpha)}/r_{\rm e(r6500)}$ in our data, and in particular the lack of significant mass or redshift dependence, we assume a constant value in the ratio of star-forming to stellar size $r_{\rm e(SF)}/r_{\rm e(M_*)}$, which we call the

size growth factor. A galaxy starts with an exponential profile at high redshift. This profile is evolved self-consistently over many small steps in time. At each step, the newly formed stars are generated with a radial distribution described by an exponential profile and a half-mass size equal to the current stellar half-mass size multiplied by the size growth factor. Stellar mass loss is self-consistently tracked as a function of the stellar age, such that mass is removed from the radii at which it was added when those stars formed, i.e., stars are assumed to remain on their initial, circular orbits. The profile evolves in this way, driving growth in the stellar disk with time. The evolving profile retains a roughly exponential form but with slowly changing scale length with radius ($n_{\rm Sersic} - 1 > 0$ but is small) such that the inner ($r \ll r_{\rm e}$) profile, dominated by old stars, is consistent with an exponential with half-light radius $= r_{{\rm e}(M_*)}$ and the outer ($r \gg r_{\rm e}$) profile, dominated by young stars, is consistent with an exponential with half-light radius equal to $r_{{\rm e}(M_*)} \times (r_{\rm e(SF)}/r_{{\rm e}(M_*)})$.

Figure 12 shows the toy model evolution of a galaxy with size growth factor $\frac{r_{\rm e(SF)}}{r_{{\rm e}(M_*)}} = 1.26$ and final stellar mass $\log_{10}(M_*/M_\odot) = 11$ at $z = 1$. Recall that dust considerations in Section 6.4 indicate that $\frac{r_{\rm e(SF)}}{r_{{\rm e}(M_*)}} = 1.26$ should be an upper limit. We examine the evolution in stellar mass and size (focused on redshifts $1 < z < 2$) and in the stellar mass versus size plane.[28] A simple empirical estimate suggests that the fractional mass evolution goes as the specific SFR, while the fractional radial evolution goes as the specific SFR times the size growth factor (ignoring mass loss and the mildly nonexponential nature of evolving profiles). With this recipe, the rate of change of log size with respect to the change in log mass is the natural logarithm of the size growth factor, i.e., $\Delta(\log r_{\rm e}) \sim \ln\left(\frac{r_{\rm e(SF)}}{r_{{\rm e}(M_*)}}\right) \times \Delta(\log M_*)$ (blue line in Figure 12). This approximation is close to our numerically derived slope, especially in the no-mass-loss case (age-dependent mass loss leads to mildly nonlinear effects).

In the bottom right panel of Figure 12 we examine how the toy model evolution with different size growth factors compares to the measured evolution of the size–mass relation for late-type (star-forming) galaxies, from vdW14, interpolated between the tabulated midpoints of redshift bins at $z = 1$ and $z = 2$, and corrected to rest-frame 6500 Å for consistency with our observed sample. This choice of observed relation is discussed further in Section 7.3.

Our evolutionary tracks are fixed such that they result in a galaxy of $\log_{10}(M_*/M_\odot) = 11$ at $z = 1$ and with a size defined by the $z = 1$ size–mass relation of vdW14. Evolving up to this fixed point, we find that a size growth factor close to our upper limit value of 1.26 is required merely to evolve galaxies *along* the size–mass relation. With smaller values, the sizes of star-forming galaxies at fixed stellar mass would actually *decrease* with increasing cosmic time. These results suggest that growth via star formation is unlikely to grow galaxies enough in size to do more than retain a nonevolving size–mass relation and almost certainly does not explain such a strong evolution as observed by vdW14.

In Figure 13 we study the ratio of $z = 1$ to $z = 2$ size as a function of the size growth factor for galaxies resulting in a galaxy of a particular mass at $z = 1$ and with a size defined by the $z = 1$ size–mass relation of vdW14. The arrows point to the location where the model size growth matches the observed growth for the average star-forming galaxy. We examine four final ($z = 1$) galaxy masses $10 \leqslant \log_{10}(M_*/M_\odot) \leqslant 11.5$. Although the progenitor galaxies at $z = 2$ would not make our KMOS^{3D} sample for the lower two mass bins (Table 2), it is still useful to examine predictions from our model under the assumption that the mass and redshift independence of the size growth factor extends to lower mass. The two vertical lines correspond to the median (1.19) and mean (1.26) values of $r_{{\rm e}({\rm H}\alpha)}/r_{{\rm e}(r6500)}$ from our analysis, respectively.

In all mass bins considered a very large size growth factor is required for galaxies to grow sufficiently to evolve from the $z = 2$ size–mass relation onto the $z = 1$ size–mass relation, assuming the vdW14 late-type relations. This also increases with mass, from $r_{\rm e(SF)}/r_{{\rm e}(M_*)} \sim 1.50$ at $\log_{10}(M_*/M_\odot) = 10$ to $\sim$1.60 at $\log_{10}(M_*/M_\odot) = 11.5$.

7.3. Considerations and Constraints on Evolution

To understand our constraints, we should consider a couple of apparent contradictions and how to resolve them.

1. vdW14 determine an evolution in late-type galaxy sizes at 5000 Å at fixed stellar mass of $\sim H(z)^{-2/3}$ and an (almost unevolving) dependence on stellar mass of $M_*^{0.22}$. Suess et al. (2019) find little dependence on stellar mass or redshift in *half-mass* sizes, accounting for gradients in the stellar mass-to-light ratio. Little redshift- or mass-dependent evolution in sizes indicates little or no growth in size.
2. Without star-formation-driven size growth, there would be no age gradients in galaxies, contradicting the idea of larger sizes at shorter wavelengths owing to the increasing importance of younger stellar populations. Instead, our results *do* support the idea that newly forming stars populate a slightly larger disk than older stars.
3. A constant size growth factor $\frac{r_{\rm e(SF)}}{r_{{\rm e}(M_*)}}$ as a function of size, mass, and redshift would imply a scale-free growth such that while younger stars can be found on average at larger galactocentric distances than older stars, the ratio of the two is independent of redshift and galaxy mass: we do not detect significant deviations from this constant growth. If true, there should be no epoch at which age gradients should disappear without invoking complex age-dependent radial migration. This appears to contradict the lack of M/L gradients at $z = 2$ seen by Suess et al. (2019).

   We do also find a greater difference between mass and light sizes in galaxies with higher Sérsic index, but only at $z \sim 1$ (not $z \sim 2$). This suggests a role for bulges, more prevalent at lower redshift, and driving greater M/L gradients in some galaxies. We note that in general star-forming disks may grow while overall galaxy sizes stay the same owing to an increasing bulge contribution: due to the differing contributions of bulge and disk to light and mass, the bulge is more likely to dominate in mass, while more size growth may be seen in light.
4. On the other hand, in Figure 9 we found no correlation of $r_{{\rm e}({\rm H}\alpha)}/r_{{\rm e}(r6500)}$ with $r_{{\rm e}(M_*)}/r_{{\rm e}(r6500)}$. This strongly suggests that whatever drives the gradients in M/L (such as age and bulge contributions) is not the dominant factor driving variations in $r_{{\rm e}({\rm H}\alpha)}/r_{{\rm e}(r6500)}$. We argue that these are driven instead mainly by variations in the gradient of dust and in particular embedded dust in star-forming disks. Continuum

[28] For simplicity in this section the simple notation $r_{\rm e}$ is frequently used to refer to the half-mass size $r_{{\rm e}(M_*)}$.

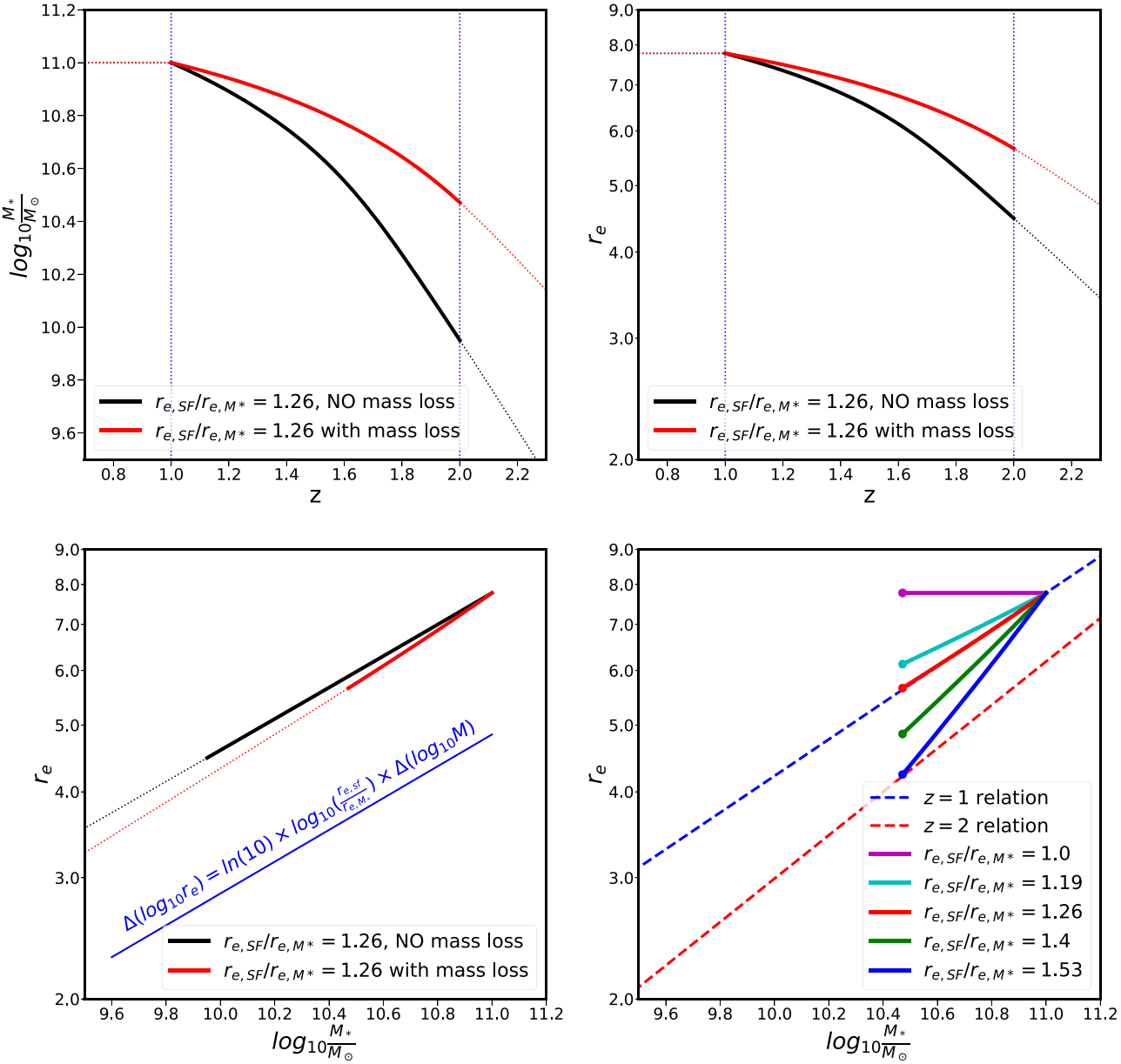

Figure 12. Toy model evolution of a galaxy with size growth factor $r_{e(SF)}/r_{e(M_*)} = 1.26$ in stellar mass (top left panel), stellar size (top right panel), and the evolutionary track in stellar mass vs. size (bottom left panel). Tracks are shown for models with (red) and without (black) mass loss. Thicker, solid lines refer to the evolution between $z = 2$ and $z = 1$. Including mass loss significantly reduces the rates of growth in both mass and size, without much changing the evolutionary track in mass vs. size. A good analytic approximation for the size evolution is shown in the bottom left panel (blue line). Bottom right panel: toy model growth from $z = 2$ to $z = 1$ as in the previous panel, but for different values of the size growth factor $r_{e(SF)}/r_{e(M_*)}$. Galaxies are evolved using our toy model and end by design on the $z = 1$ size–mass relation of star-forming (LT) galaxies from vdW14 corrected from F160W to rest-frame 6500 Å using the Kelvin et al. (2012) recipe (blue dashed line). The evolution of galaxy size depends on the size growth factor $r_{e(SF)}/r_{e(M_*)}$ as depicted by the solid evolutionary track lines. A very large size growth factor $r_{e(SF)}/r_{e(M_*)} \sim 1.53$ is required to reproduce the observed evolution, from the $z = 2$ LT size–mass relation of vdW14 (red dashed line), while if we assume size growth factors equivalent to our measured mean (median) observed values $r_{e(H\alpha)}/r_{e(F160W)} = 1.26\,(1.19)$, these are barely enough to evolve a galaxy *along* a nonevolving size–mass relation.

light and Hα are therefore more closely tied than mass and Hα, and the ratio should remove the effect of a foreground dust screen but not of embedded dust.

Taking account of these considerations, we selected to compare to the simpler, light-based vdW14 relations. Considering our results, we cannot match the observed evolution of size–mass relations from vdW14. This would suggest that other physical processes for growth of star-forming galaxies might be at play, and we will discuss candidates in the next section. However, the results of Suess et al. (2019) suggest that evolution might not be so steep if we assume that there is also a role for evolution in the mass-to-light gradients (even if this is driven in part by bulge formation). Such milder evolution can be consistent with our upper limit of $\sim$26% for the size growth factor.

7.4. Physical Origins of Galaxy Size Growth

Equation (6) conveniently separates the dependencies on redshift and mass in the derivation of disk size for star-forming galaxies. In this context the mass dependence is simply an imprint

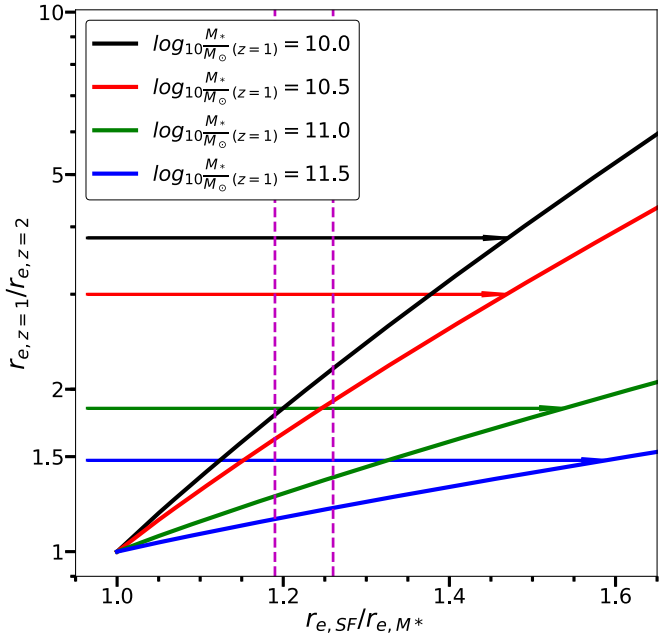

Figure 13. Total size growth in our toy model from $z = 2$ to $z = 1$ for galaxies with four values of final stellar mass at $z = 1$ (colored lines; see legend). Sizes are assumed to grow exclusively via the formation of new stars with a constant size growth factor $r_{e(SF)}/r_{e(M_*)}$ (x-axis), and with a mass evolution following the star formation MS. The arrows denote the equivalent size growth of the same galaxy over the same redshift interval from the size–mass relation of vdW14 and point to the place where this growth is reached by our size growth model. These values are much larger than those measured in our sample (vertical, dashed magenta lines show the median and mean values at 1.19 and 1.26). This suggests that star-formation-driven size growth is not enough to explain the observed size evolution of star-forming galaxies.

of the dependence of galaxy size on $\left(\frac{M_d}{m_d}\right)^{1/3} = M_{\rm halo}^{1/3} \propto V_c$ and thus an imprint of the Tully–Fisher relation between galaxy mass and circular velocity. As star formation appears to drive evolution approximately along this relation, this implies that the integrated growth in size and mass can be described by Equation (6), with $R_d \propto (M_d/m_d)^{1/3}$ and thus with $M_{\rm halo}^{1/3}$, but with galaxy sizes scaled to the Hubble parameter at the epoch of observation, $H(z_{\rm obs})$, such that there is no redshift dependence of the size growth factor (Equation (7)). The measured size evolution of the size–mass relation by vdW14, on the other hand, scales as $H(z_{\rm obs})^{-2/3}$ (or alternatively as $(1+z)^{0.75}$). Although this evolution might be overestimated in light-weighted sizes (see Section 7.3), if real it would suggest another form of growth not associated with star formation—i.e., that the stellar component of star-forming galaxies does not retain its initial size, but rather evolves in size as predicted by Equation (6) owing to angular momentum transfer with the surrounding material (gas and dark matter; see, e.g., Struck & Elmegreen 2017). In addition, as we discuss below, *any* measured evolution of the size–mass relation is likely to be an overestimate, given that many of the more compact massive galaxies have their star formation quenched between epochs.

The constant size growth factor, with no obvious dependence on redshift, size, and stellar or gas mass, implies that the model star-formation-driven size growth as seen in Figures 12 and 13 applies under widely varying conditions. While this may simply reflect the halo mass growth, the small intrinsic scatter (43%) also implies that the halo spin parameter λ remains very stable over time and that the specific angular momentum transfer ($\frac{\lambda'}{\lambda} = \frac{j_d}{m_d}$ in Equation (6)) is also quite insensitive to a wide variety of halo growth rates and physical conditions. Such stability can be achieved if disk growth is regulated, e.g., via feedback. For example, Pezzulli et al. (2017) describe gas accretion from a rotating hot corona gas in the halo. The accretion of such gas—with its associated angular momentum—is expected in a galactic fountain model in which stellar winds interact with the corona gas before falling back onto the disk. Such models explain the rotation lag of extraplanar gas in the Milky Way galaxy (Marinacci et al. 2011) and can help explain a slow growth in the size of the star-forming disk, strictly linked to the stellar disk size with small scatter.

Our results and modeling suggest that individual galaxies evolve almost parallel to the size–mass relation with a maximum evolution at $\frac{d\log(r_e)}{d\log(M_*)} \sim 0.26$ for $r_{e(SF)}/r_{e(M_*)} = 1.26$ (Figure 12), in agreement with the models from Nelson et al. (2019). The robustness of our measurement is also supported by comparisons to a completely independent estimate coming from expectations when comparing the sizes of Milky Way progenitor galaxies selected at different redshifts assuming a constant cumulative comoving number density ($\frac{d\log(r_e)}{d\log(M_*)} \sim 0.27$ from van Dokkum et al. 2013), and it is slightly shallower than the slope of ∼0.3 discussed by van Dokkum et al. (2015) and ∼0.4 for simulated galaxies with realistic wind models (Hirschmann et al. 2013). In contrast to models with no winds, the efficient removal of low angular momentum material at high redshift leads to much larger sizes for high-redshift galaxies and shallower evolution, emphasizing the role of feedback in the regulation of angular momentum in galaxies.

Some of the massive, star-forming $z = 2$ galaxies will have had their star formation quenched by the time they reach $z = 1$. There are different theories of how such quenching proceeds and why it returns passive galaxies more compact than the coeval star-forming population: galaxies with low ($\lambda \lesssim 0.05$) spin parameters can become unstable and contract before rapidly being quenched (Dekel & Burkert 2014); galaxies reach a threshold stellar surface density, velocity dispersion, central surface density, stellar mass, or bulge-to-total ratio before quenching (e.g., van Dokkum et al. 2015), or galaxies evolve along the MS and the effective quenching of massive galaxies is a gradual process, with older, earlier-forming galaxies with higher density and smaller sizes departing first from the MS (Lilly & Carollo 2016; Abramson & Morishita 2018). Abramson & Morishita (2018) argue that the distribution of galaxies in the size–mass plane is not inconsistent with such a scenario in which galaxies evolve at constant surface mass density $\left(\frac{d\log(r_e)}{d\log(M_*)} = 0.5\right)$. Our constraints show that such a steep evolution is only possible if the stellar sizes of galaxies grow via mechanisms other than star formation. As shown—for very different models—by van Dokkum et al. (2015) and Abramson & Morishita (2018), such a strong apparent evolution can happen even if the evolution of individual galaxies is relatively weak, so long as the densest galaxies fall out of the star-forming population first and become passive. Our weaker measured star-formation-driven evolution in size versus mass suggests that a more aggressive quenching is required, resulting in passive galaxies that are particularly dense and compact.

8. Summary and Conclusions

This paper utilizes data from the KMOS^{3D} survey to measure the star-formation-driven size growth in individual galaxies and understand the physical processes driving this evolution at $0.7 \lesssim z \lesssim 2.7$, spanning the time when most of their stars were formed.

KMOS^{3D} targeted the Hα+[N II] emission-line complex in 740 galaxies at $0.7 \lesssim z \lesssim 2.7$ with 75 nights of observation using the multiplexing NIR IFU instrument KMOS on the VLT. Data cubes and associated bootstrap cubes are released with an associated data release paper (W19). In this paper we derive galaxy sizes in Hα emission, tracing ongoing star formation. Our initial goal was to demonstrate the accurate measurement of galaxy half-light sizes with ground-based data at these redshifts, and in particular Hα sizes tracing star formation. With the investment of a significant calibration effort we have achieved our goal, with galaxy sizes measured with a typical accuracy of ~20%, almost independent of the absolute size. Our analysis resulted in a sample of 281 galaxies for which we have accurate Hα size measurements with associated errors. It is representative of the overall star-forming population in terms of SFR, stellar mass, colors, and continuum size. We publish sizes and associated errors for this sample.

We then examined how the size of the star-forming gas, traced by Hα, relates to other galaxy properties. Our results can be summarized as follows:

1. Hα sizes depend primarily on the continuum size of a galaxy, with a near-linear relation, a median (mean) Hα size $= r_{e(H\alpha)}$ 18% ± 3% (~26%) larger than $r_{e(r6500)}$, and the continuum size at rest-frame 6500 Å, with 43% ± 3% intrinsic lognormal scatter. This is much tighter than other correlations (e.g., 68% ± 4% intrinsic scatter in Hα size vs. stellar mass) and explains most of the variation in Hα size. It is also smaller than the intrinsic scatter in continuum size versus stellar mass (~56%).
2. The dependence of Hα size on continuum size shows no residual dependence on stellar mass, redshift, star formation activity, galaxy morphology, or (indirectly) versus gas mass.
3. There is a significant residual dependence of Hα size on dust extinction properties affecting the Hα emission. This dependence does not arise from the continuum extinction A_V. Instead, the size ratio depends primarily on the amount of extinction in the stellar birth clouds $A_{H\alpha}$ compared to A_V. For galaxies with larger Hα disks the ratio of these two extinction measurements tends to be lower, closer to a pure foreground screen approximation. This is in line with models suggesting that most obscuration of Hα emission far from galactic centers takes place in a diffuse, foreground component (equally affecting the continuum), while in galactic centers more obscuration of Hα is caused by dust embedded in H II regions (e.g., Li et al. 2019).

Based on these results, we surmise the following:

1. The tight correlation between the sizes of the star-forming and stellar components in star-forming, high-redshift galaxies infers that the spin parameters of galaxies and their gas are tightly linked to the halo in which they live (see also Burkert et al. 2016) and are highly stable over long periods of cosmic time. Such stability requires not only stable halo growth but also stability in the transfer of specific angular momentum from halo to disk scales, including the processes of cooling, accretion, star formation, and feedback. Simulations and models suggest that feedback helps to modulate the disk growth (e.g., Hirschmann et al. 2013) and might help regulate the growth and its relation to the existing stars via galactic fountains (e.g., Pezzulli et al. 2017). Such regulation applies consistently across a wide range of physical conditions as characterized by, e.g., redshift, stellar or gas mass, SFRs, morphology, environment, and global extinction by dust, A_V.
2. Star formation drives size growth in galaxies but is unlikely to evolve galaxies more steeply in size versus mass than the observed relation between those parameters at fixed redshift. We model this process including stellar mass loss. Excess dust extinction of Hα in galaxy centers means that our mean value of $r_{e(H\alpha)}/r_{e(F160W)} = 1.26$ is likely an upper limit on the size growth factor $r_{e(SF)}/r_{e(M_*)}$. A growth of $r_{e(SF)}/r_{e(M_*)} = 1.26$ moves galaxies along a locus with a slope $\frac{d\log(r_e)}{d\log(M_*)} \sim 0.26$, consistent with the slope of the mass–size relation and with the predicted evolution of Milky Way progenitors based on the observed sizes of galaxies selected at a constant cumulative comoving number density in the universe ($\frac{d\log(r_e)}{d\log(M_*)} \sim 0.27$; van Dokkum et al. 2013). A steeper evolution of the sizes of star-forming galaxies with mass, required to explain the roughly $H(z)^{-2/3}$ evolution in stellar sizes (vdW14), may be partially accommodated by accounting for the quenching of star formation in the more compact galaxies (e.g., van Dokkum et al. 2015; Abramson & Morishita 2018) but likely requires other physical processes such as minor merging and angular momentum exchange with the halo.

D.J.W. and M.F. acknowledge the support of the Deutsche Forschungsgemeinschaft via Projects WI 3871/1-1, and WI 3871/1–2. M.F. has received funding from the European Research Council (ERC) under the European Union's Horizon 2020 research and innovation program (grant agreement No 757535). E.W. and J.T.M. acknowledge support by the Australian Research Council Centre of Excellence for All Sky Astrophysics in 3 Dimensions (ASTRO 3D), through project No. CE170100013. P.L. acknowledges funding from the European Research Council (ERC) under the European Union's Horizon 2020 research and innovation program (grant agreement No. 694343). G.B. acknowledges support from the Cosmic Dawn Center, which is funded by the Danish National Research Foundation. We thank the ESO and Paranal staff for the excellent support over the course of our KMOS observations. This work is also based on observations taken by the 3D-*HST* Treasury Program (GO 12177 and 12328) and by the CANDELS Multi-Cycle Treasury Program with the NASA/ESA *HST*, which is operated by the Association of Universities for Research in Astronomy, Inc., under NASA contract NAS5-26555.

Appendix A
Size Measurements: Accuracy and Consistency

The main goal of this paper is to present measurements of the size in Hα for KMOS^{3D} galaxies. With complimentary

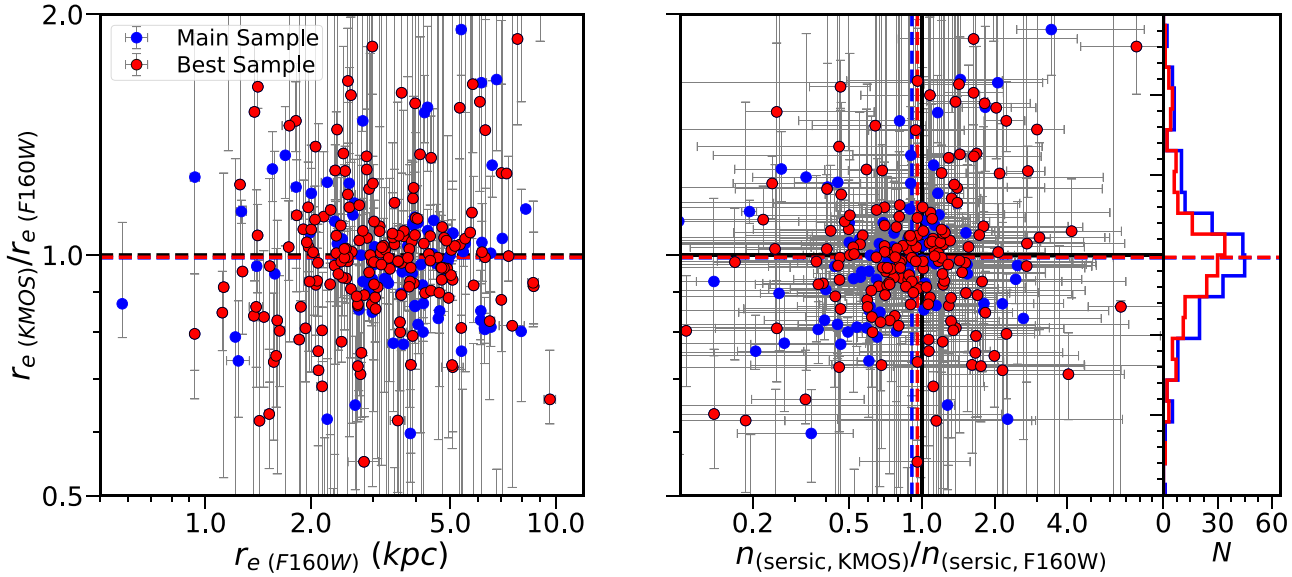

Figure 14. Left: ratio of KMOS continuum-based effective radius (r_e) to that from CANDELS (F160W), as a function of the CANDELS F160W r_e for galaxies in the continuum MAIN (blue) and BEST (red) samples, and its distribution (narrow panel to the right). Errors on the size ratio are derived from bootstrap errors on KMOS continuum sizes. The median ratio is 0.99 for both samples (dashed horizontal lines) and does not notably depend on the galaxy size, demonstrating the well-recovered continuum sizes of galaxies from KMOS data. Right: comparison of the ratio of parameters—size (r_e) and Sérsic index ($n_{\rm Sersic}$) —from fitting to the KMOS continuum data to those from fitting the CANDELS F160W-band images, with bootstrap error bars and their medians (horizontal and vertical dashed lines). Both parameters are well reproduced with fits to the KMOS continuum data. Galaxies are divided into the continuum MAIN and BEST samples.

continuum measurements this provides the means to examine the spatial growth of high-redshift galaxies via star formation. This section provides a characterization and tests of the accuracy of our measurements with natural seeing KMOS data for both continuum and Hα.

In Figure 3 we have seen that our IMFIT-based sizes measured with the F160W WFC3 band are consistent with those measured by vdW14. Moreover, the right panel shows a remarkable consistency between our higher-resolution WFC3 sizes and those derived from the KMOS continuum.

A.1. Definition of Galaxy Size Errors

The accuracy of measured half-light size of galaxies is sensitive to sources of noise (systematic and random) and to the spatial resolution.

Our bootstrap cubes randomly sample most sources of systematic and random noise, so we use these to assess asymmetric 1σ errors in size. Bootstrap size errors encompass the range between the 16th and 84th percentiles of the sizes measured from the bootstrap cubes.

In some cases the median size measured from the bootstrap cubes differs from the size measured using our total combine cube. To be conservative, we also define statistical size errors such that the negative and positive errors are each the maximum of the difference between the median bootstrap or total combined, and the 16th and 84th bootstrap percentiles, respectively.

Sizes based on the bootstrap cubes do not account for any uncertainty on the PSF. Therefore, in Section A.4 we assess the impact of PSF uncertainty on galaxy size measurements, assigning a minimum size error based on the uncertainty due to the PSF. Our final size errors are the maximum of the statistical errors, and this minimum error is based on the PSF uncertainty.

In the following discussion we shall examine our bootstrap, statistical, and final errors. Our final results use final errors in all cases.

A.2. Errors on Galaxy Sizes: Continuum

We now examine the accuracy of KMOS continuum sizes, assessed via comparison with the higher-resolution sizes measured on CANDELS WFC3 data. In the left panel of Figure 14, we find that the galaxy sizes fit to the KMOS continuum image are equivalent to those from fits to the higher-resolution and higher-S/N CANDELS F160W-band images, with a median offset of just 1% for both continuum MAIN and BEST samples and no apparent dependence on galaxy size. The right panel extends this to the joint parameter space of size and Sérsic index: Sérsic indices are also compatible, with median offsets of $<10\%$.

Figure 15 examines the distribution of these size offsets compared to our derived errors. In the top left panel we show the cumulative distribution of size offsets normalized by the size itself $\left(\frac{\rm r_e(KMOS)-r_e(F160W)}{\rm r_e(F160W)}\right)$ and of our estimated size errors, also normalized by size. We measure more small fractional offsets in size than would be predicted by our measurement errors, suggesting that some errors are slightly overestimated. This is mostly due to the difference between bootstrap and statistical errors—i.e., to account for differences between the median bootstrap realization and the best estimate. The top right panel examines this fractional (final) error and offset distribution separately for the most compact galaxies ($r_e < 2$ kpc) and for more extended galaxies, demonstrating little difference in the accuracy of sizes or of size errors, perhaps because compact sources tend to be brighter, compensating for the lack of resolution with higher S/N. In the bottom left panel the size offsets are normalized by the size error. Especially using our final errors, this describes something very close to an error function with a dispersion of 1 and median of 0, as would be expected in the case that the errors are accurate. Therefore, we consider our size errors (at least for KMOS continuum sizes) to be well calibrated.

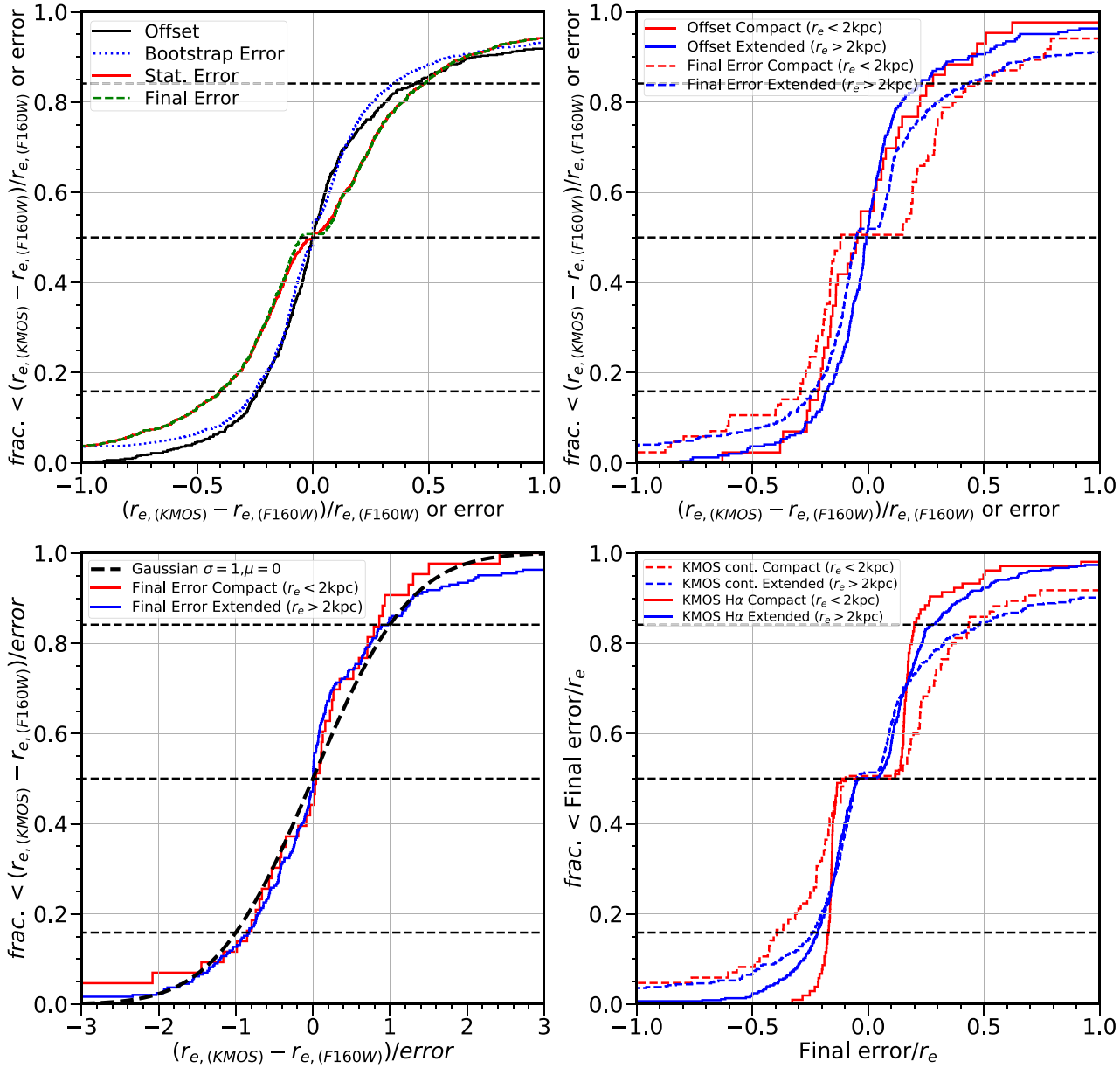

Figure 15. Top left: cumulative distribution of the relative difference (offset) between KMOS continuum and CANDELS F160W based size measurements $\left(\frac{r_e(\mathrm{KMOS}) - r_e(\mathrm{F160W})}{r_e(\mathrm{F160W})}\right.$; solid black line$\left.\right)$ from the continuum MAIN sample. This is compared to the cumulative distribution of the expected size error distribution for these galaxies: we show bootstrap (dotted), statistical (dashed), and final (solid) error distributions (see Section A.1 for error definitions). Top right: offsets and final error cumulative distributions, divided into compact (red) and extended (blue) sources at $r_e = 2$ kpc. Remarkably, fractional offsets are similar for compact and extended galaxies, and while the final error (which accounts for PSF uncertainty) inevitably gives larger fractional errors for compact galaxies, this converges to similar values for larger fractional errors. Bottom left: cumulative distribution of the error-normalized size offset: $\frac{r_e(\mathrm{KMOS}) - r_e(\mathrm{F160W})}{\sigma(r_e(KMOS) - r_e(F160W))}$ again divided into compact and extended sources at $r_e = 2$ kpc. Galaxies offset negatively from their F160W sizes are normalized by the negative error and vice versa. With accurate errors, this should describe a normal distribution with a mean of 0 and scatter of 1, the cumulative version of which is the equivalent error function (black dashed line). This provides a remarkably good match to both compact and extended sources, indicating that our final errors are accurate, not a function of size, and applicable to galaxies sampled with KMOS resolution with our best-guess KMOS PSF. Bottom right: cumulative distribution of fractional (final) size errors for KMOS continuum and Hα, selected respectively from the continuum MAIN and Hα MAIN samples, and divided into compact and extended galaxies at $r_e = 2$ kpc (independently for KMOS continuum and Hα measurements). While the relative errors on KMOS continuum sizes are slightly larger for compact galaxies, the opposite is true for the Hα case: compact galaxies have smaller relative errors on average.

A.3. Errors on Galaxy Sizes: Hα

The bottom right panel of Figure 15 now examines the errors for both KMOS continuum sizes and KMOS-based Hα sizes. This shows the cumulative distribution of fractional errors on KMOS continuum and Hα sizes from the respective MAIN samples and divided into compact and extended subsamples (at 2 kpc).

While the error distribution for KMOS continuum and Hα is very similar for extended sources, the compact Hα galaxies have significantly smaller errors than the compact KMOS continuum galaxies.

We have shown that the errors on the KMOS continuum sizes are well described by (systematic and random) S/N variations as traced primarily by the bootstrap errors. Figure 15

demonstrates that the compact (typically high-S/N) Hα sources can have very small size errors. This motivates a more thorough examination of the effects of uncertainty on the PSF in Section A.4, resulting in the final size errors. We note here, based on a comparison of bootstrap/statistical and final errors in Figure 15, that this inflates the smallest size errors for compact galaxies, but that there are nonetheless fewer compact galaxies with large fractional size errors in Hα than for extended galaxies or for KMOS continuum, implying that this is robust and likely a consequence of the high-S/N data. We also tested the impact of our choice of a flat extension of the velocity field beyond the regions where we trust kinematic fits in individual spaxels. We modified the velocity in these extrapolated regions to values 25% above and below the nominal values. This simulates either declining rotation curves at large radii (Lang et al. 2017) or rotation curves that keep rising to larger velocities. The impact on the final Hα sizes is negligible, with a scatter with respect to the best values of 1%–2% in both our tests. This scatter is much smaller than the individual measurement errors, which are therefore not affected by the algorithm used to extend the velocity fields.

A.4. Effects of PSF Uncertainty on Size Errors

As described in Section 3.2, the generation of PSF images for each combined cube relies on our ability to accurately measure the shifts between exposures, acquisitions, and setups, as characterized via the simultaneously observed stars, and shifts between partial combines. An uncertainty on the PSF (e.g., from arm positioning errors not accounted for via shifts between partial combines) translates to a limit to our effective size or size error estimates that is not accounted for by the bootstrap errors, but which must be smaller than the typical error on the KMOS continuum size (given that those bootstrap errors are large enough to describe the offset from CANDELS based sizes; Section A.2).

To examine how the size estimates are sensitive to the assumed PSF, we refit every galaxy in the sample with the PSF (image) as computed for all of the other galaxies in the sample. This covers a much broader range of PSF than any realistic, residual error on the true PSF. Figure 16 shows the resultant variation of best-fit galaxy half-light radius with assumed PSF (circularized) FWHM for one of our compact galaxies, COS4_09044. The dependence on the size of the PSF is most closely related to the circularized size, $r_{e,circ}$, expressed in angular units (arcseconds).[29]

Focusing on the results for the fit to the KMOS continuum (black points), we see the expected trend as typically seen for most galaxies (including more extended ones): as a larger PSF is assumed, the fitting procedure compensates such that the best-fit intrinsic galaxy size is more compact. To illustrate the range of assumed PSF that can be accommodated within the 1σ size range derived using the statistical errors from the bootstrap cubes (dashed black horizontal lines)—i.e., the range of PSF for which the statistical errors dominate over any size error induced by the assumption of the incorrect PSF—we fit the dependence of predicted galaxy size on assumed PSF FWHM with a fifth-order polynomial (green solid line, fit is iterative with sigma clipping) and determine the range of PSF FWHM for which this fit lies within the statistical errors on size (dotted vertical black lines), with a full range $2 \times \Delta\mathrm{FWHM}_{\mathrm{PSF,circ}}$.

[29] Circularized sizes are defined as $r_{e,circ} = r_e.\sqrt{1-\epsilon}$, $\mathrm{FWHM}_{\mathrm{PSF,circ}} = \mathrm{FWHM}_{\mathrm{PSF,major}}.\sqrt{1-\epsilon_{\mathrm{PSF}}}$.

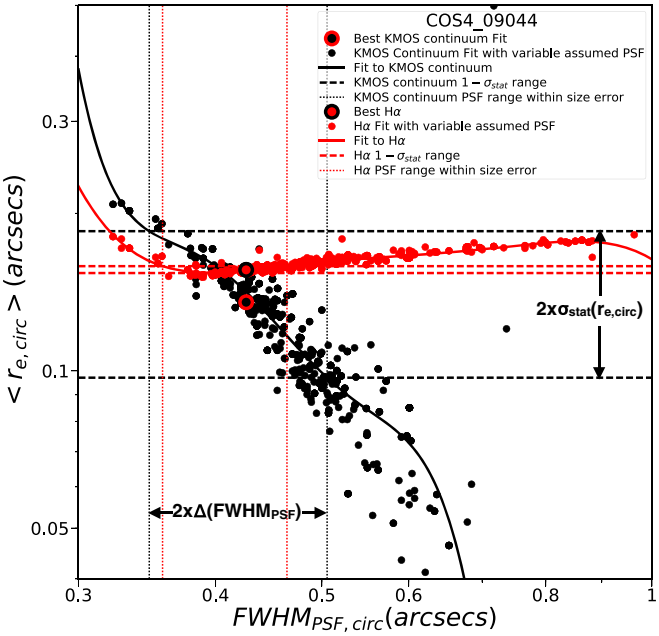

Figure 16. Best-fit circularized effective radius, $< r_{e,circ} >$ expressed in angular units, for compact galaxy COS4_09044 in KMOS continuum (black) and Hα (red) as a function of the FWHM of the circularized PSF ($\mathrm{FWHM}_{\mathrm{PSF,circ}}$), which is assumed for the fit (to the same data). The fit with the best-guess PSF for this galaxy is indicated with the larger, outlined data points. With a larger PSF, the best-fit size to KMOS continuum data decreases to compensate. This is representative of the typical behavior for most galaxies. Only if the assumed PSF is well away from our best guess will the fit to the KMOS continuum data provide a best-fit galaxy size outside the range allowed by the statistical errors from bootstrap cubes (dashed horizontal black lines), which samples random and systematic variations in the input data. We fit the trend with a fifth-order polynomial including sigma clipping (black solid line) and measure the range of assumed $\mathrm{FWHM}_{\mathrm{PSF,circ}}$ accommodated within the statistical errors (black dotted vertical lines, the separation of which is defined to be $2 \times \Delta\mathrm{FWHM}_{\mathrm{PSF,circ}}$). The Hα fits, in contrast, provide a very stable circularized galaxy size almost independent of bootstrap iteration (red dashed horizontal lines) *or of the assumed PSF* (lack of variation with $\mathrm{FWHM}_{\mathrm{PSF,circ}}$, fit indicated as the red solid line). This is typical of high-S/N Hα data for compact galaxies (r_e(Hα) $\lesssim 0\farcs25$ or 1.25 pixels) as seen in Figure 17.

In Figure 17 we show how $\Delta\mathrm{FWHM}_{\mathrm{PSF,circ}}$ depends on the statistical error range of the galaxy size ($\sigma_{\mathrm{stat}}(r_{e,circ})$ in arcseconds, the average of positive and negative errors). $\Delta\mathrm{FWHM}_{\mathrm{PSF,circ}}$ saturates at a maximum value, corresponding to the full range of $\mathrm{FWHM}_{\mathrm{PSF,circ}}$. For smaller size errors, there exists some range of assumed PSF that would drive the galaxy size outside the statistical error: the range of PSF consistent with the measured size and statistical errors becomes smaller with decreasing statistical error along a locus of slope ~0.9 (a power law on linear scales with exponent 0.9). Repeating the exercise for Hα fits (red points in Figures 16 and 17), we see that for most cases (in particular, for galaxy sizes $r_e \gtrsim 0\farcs25$) the data follow the same trend, for which we perform a linear fit (green solid line in Figure 17).

We estimate that there can exist a *maximum* residual error of ~1 pixel $= 0\farcs2$ on $\mathrm{FWHM}_{\mathrm{PSF,circ}}$, as a result of uncertain manual shifts between partial combines, and residual errors after the average exposure to exposure shift of PSF stars has been removed. This sets a conservative upper limit on the error on $\mathrm{FWHM}_{\mathrm{PSF,circ}}$ for a compact PSF ($\mathrm{FWHM}_{\mathrm{PSF,circ}} \sim 0\farcs4$) of $\sim\sqrt{(0\farcs4)^2 + (0\farcs2)^2} - 0\farcs4 \sim 0\farcs05$. We assume that where $\Delta\mathrm{FWHM}_{\mathrm{PSF,circ}} < 0\farcs05$ (horizontal dashed black line,

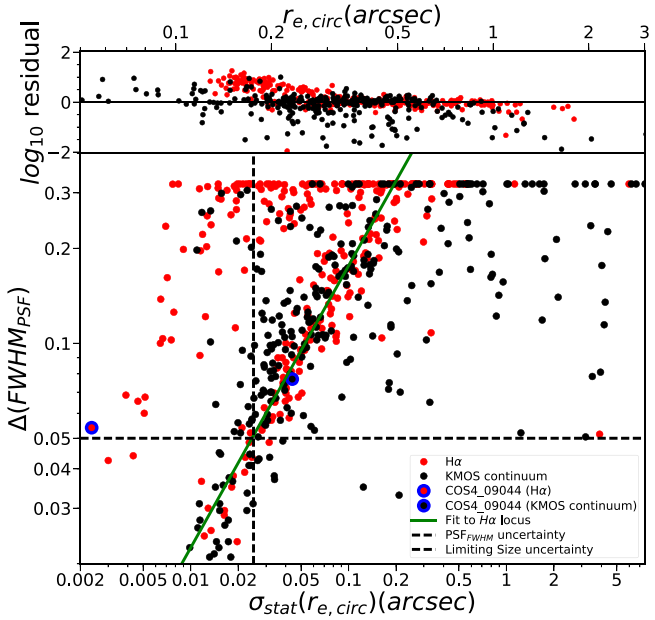

Figure 17. Each galaxy is fit with the full range of assumed PSF, as described in Section A and illustrated by Figure 16. The *statistical* error on galaxy size, $\sigma_{\rm stat}(r_{\rm e})$, is estimated using the fits to bootstrap cubes. Here the statistical error on circularized galaxy size, $\sigma_{\rm stat}(r_{\rm e,circ})$, is shown plotted aginst the half-range of PSF circularized FWHM, $\Delta{\rm FWHM}_{\rm PSF,circ}$, for which the best-fit galaxy size lies within the statistical error (i.e., the range of PSF error for which the statistical error dominates the error on the assumed PSF). All galaxies in the continuum MAIN (black, KMOS continuum) and Hα MAIN (red) are shown. The vast majority of galaxies in KMOS continuum, and many in Hα, lie along a locus of decreasing $\Delta{\rm FWHM}_{\rm PSF,circ}$ with decreasing $\sigma_{\rm stat}(r_{\rm e,circ})$ (solid green line)—such that for very small statistical size errors the error on assumed PSF can dominate. As described in the main text, we assume a conservative error on the assumed PSF of $\Delta{\rm FWHM}_{\rm PSF,circ} = 0\farcs05$ (horizontal dashed black line), for which the error is matched by a statistical error of $\sigma_{\rm stat}(r_{\rm e,circ}) = 0\farcs025$, or $\frac{1}{8}$ of a KMOS pixel (vertical dashed black line). At very compact sizes, $r_{\rm e,circ} \lesssim 2\farcs5$, the statistical error on Hα sizes becomes very small, and the sensitivity to $\Delta{\rm FWHM}_{\rm PSF,circ}$ becomes quite flat, as seen in Figure 16 for COS4_09044. These galaxies are to the left of the main locus of points in this figure and are clearly shown with positive residuals in the top panel, in which we show the residual of $\Delta{\rm FWHM}_{\rm PSF,circ}$ with respect to the Hα locus vs. galaxy size. We choose to apply a conservative lower limit to the circularized galaxy size error, $\sigma(r_{\rm e,circ}) \geqslant 0\farcs025$, corresponding to the vertical dashed line.

Figure 17) the PSF error overrides the statistical error: this sets a lower limit on the final error on the size at the point where our locus of points (green line) intersects this limiting value of $\Delta{\rm FWHM}_{\rm PSF,circ}$, such that our final error on circularized galaxy size $\sigma(r_{\rm e,circ}) \geqslant 0\farcs025$ (vertical dashed black line, $=\frac{1}{8}$ of a pixel).

For the fits to compact galaxies in Hα ($r_{\rm e} < 0\farcs25$), the statistical errors on size from the bootstraps are often very small, $\sigma_{\rm stat}(r_{\rm e,circ}) < 0\farcs01$ or <0.05 pixels, and with a much shallower dependence of best-fit size on assumed PSF (and thus larger value of $\Delta{\rm FWHM}_{\rm PSF,circ}$ at fixed $\sigma_{\rm stat}(r_{\rm e,circ})$). A good example is COS4_09044, highlighted in blue in Figure 17, for which the Hα fit provides an extremely consistent size almost independent of the bootstrap iteration or the assumed $\Delta{\rm FWHM}_{\rm PSF,circ}$ (red points in Figure 16). Such galaxies have high S/N and high surface brightness (surface brightness increases with decreasing size), but the best correlation is with size: the residuals with respect to the main locus in Figure 17 are plotted against galaxy size in the top panel. We conservatively set the errors of all such fits to our adopted minimum value of $\sigma(r_{\rm e,circ}) \geqslant 0\farcs025$ (vertical dashed black line). Most galaxies, especially in the case of KMOS continuum, have larger statistical errors, and so these error estimates remain effectively consistent with the differences between KMOS continuum and CANDELS sizes. We now define the final size errors on (major-axis) sizes, $\sigma(r_{\rm e})$, to be the maximum of the statistical error and the uncertainty on size due to the PSF uncertainty of $0\farcs025 \times \sqrt{1-\epsilon} \times D_A$, where D_A is the angular diameter distance at the redshift of the galaxy in kpc arcsec^{-1}.

Finally, we note that there are no fits to Hα data that are flagged as OK for which best-fit sizes are below $r_{\rm e} \sim 0.69$ pixels. This seems to be the real limit for galaxy Hα sizes in our sample, as there is no reason that smaller sizes should be flagged as bad (even if perfectly described by the PSF). This corresponds to a minimum physical size of $r_{\rm e} \sim 1.1$ kpc, very similar to the minimum size from CANDELS F160W continuum imaging (Figure 3). In contrast, a few KMOS continuum sizes reach to both much lower and larger sizes (as seen in Figure 3): these can be explained as outliers and are mostly consistent with the tails in the difference between KMOS continuum and CANDELS F160W sizes, normalized by the size errors, i.e., they are mostly expected given the errors, with a few possible exceptions.

Appendix B
Examples of Hα Profiles and Exponential Fits

In Figure 18 we show a gallery of Hα profiles from the MAIN sample spanning a range of size, redshift, and observed surface brightness, to show the quality of the data and of the fitting procedure.

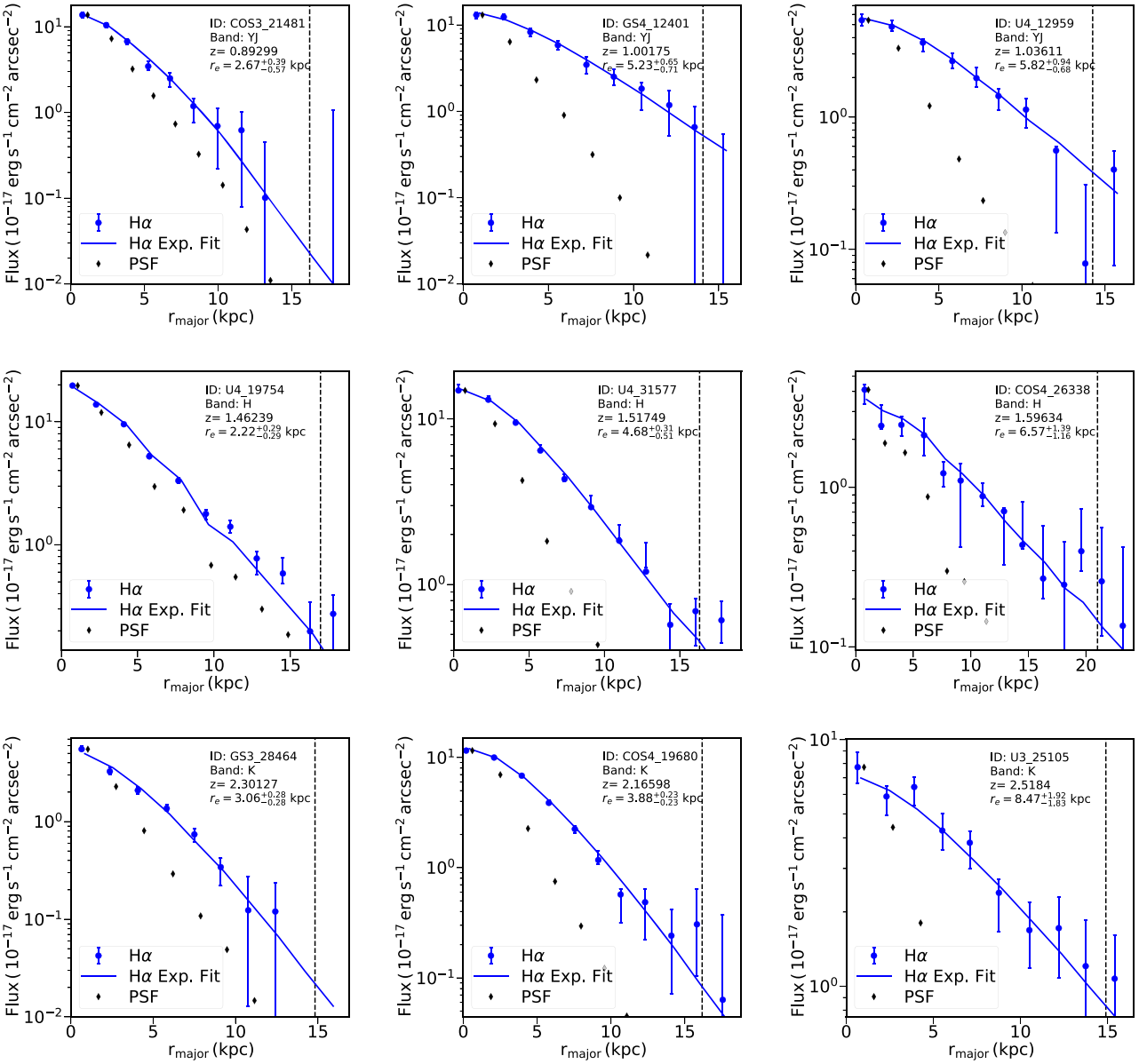

Figure 18. Gallery of radial Hα profiles extracted using elliptical apertures from our KMOS data (blue points with 1σ bootstrap errors) as described in Section 4. The blue solid lines show the 1D profiles of the best-fit 2D exponential model. For comparison the PSF image is also extracted in the same apertures (black diamonds). The vertical dashed line indicates the radius where the major axis first crosses the edge of the KMOS FOV. The galaxies have increasing effective radii from the left to the right column, while the galaxy redshift increases from top to bottom.

Appendix C
Public Release of Size Measurements

The size measurements of 281 galaxies in the main sample are made available as a machine-readable table. Table 3 gives an example of the quantities provided with this work and the description of the columns.

Table 3
Example of the Size Measurements Table Made Available with This Work

KMOS^{3D} ID	KMOS^{3D} ID TARGETED	z_{spec}	$r_{e\ (F160W)}$ (kpc)	$\sigma(r_{e\ (F160W)})$ (kpc)	$r_{e\ (H\alpha)}$ (kpc)	$\sigma_{neg.}(r_{e\ (H\alpha)})$ (kpc)	$\sigma_{pos.}(r_{e\ (H\alpha)})$ (kpc)
(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
COS4_06327	COS3_06511	0.80364	2.458	0.034	3.803	0.453	0.842
GS4_34568	GS4_34568	2.57255	3.845	0.071	5.125	0.612	0.940
U4_09733	U4_09733	2.28886	3.931	0.092	11.243	1.051	1.170
…							

Note. Column (1): object ID as defined in the data release (W19). Column (2): object ID that defined the target at the time of the observations as defined in W19. In this paper we refer to these IDs. Column (3): spectroscopic redshift based on KMOS^{3D} emission-line detection. Column (4): $r_{e(F160W)}$ effective radius from WFC3/F160W images in kpc. Column (5): $\sigma(r_{e(F160W)})$ symmetric 1σ uncertainty on the effective radius from WFC3/F160W images in kpc. Column (6): $r_{e\ (H\alpha)}$ effective radius from KMOS Hα images in kpc. The flux image is derived as described in Equation (4). Column (7): $\sigma_{neg.}(r_{e(H\alpha)})$ asymmetric negative 1σ uncertainty on the effective radius from KMOS Hα images in kpc. The final error is as defined in Appendix A. Column (8): $\sigma_{pos.}(r_{e(H\alpha)})$ asymmetric positive 1σ uncertainty on the effective radius from KMOS Hα images in kpc. The final error is as defined in Appendix A.

(This table is available in its entirety in machine-readable form.)

ORCID iDs

Matteo Fossati https://orcid.org/0000-0002-9043-8764
J. Trevor Mendel https://orcid.org/0000-0002-6327-9147
Roberto Saglia https://orcid.org/0000-0003-0378-7032
Emily Wisnioski https://orcid.org/0000-0003-1657-7878
Stijn Wuyts https://orcid.org/0000-0003-3735-1931
Natascha Förster Schreiber https://orcid.org/0000-0003-4264-3381
Alessandra Beifiori https://orcid.org/0000-0001-8017-6097
Ralf Bender https://orcid.org/0000-0001-7179-0626
Sirio Belli https://orcid.org/0000-0002-5615-6018
Hannah Übler https://orcid.org/0000-0003-4891-0794
Philipp Lang https://orcid.org/0000-0002-5681-3575
Jeffrey C. C. Chan https://orcid.org/0000-0001-6251-3125
Rebecca L. Davies https://orcid.org/0000-0002-3324-4824
Erica J. Nelson https://orcid.org/0000-0002-7524-374X
Reinhard Genzel https://orcid.org/0000-0002-2767-9653
Linda J. Tacconi https://orcid.org/0000-0002-1485-9401
Dieter Lutz https://orcid.org/0000-0003-0291-9582
Sedona Price https://orcid.org/0000-0002-0108-4176
Andreas Burkert https://orcid.org/0000-0001-6879-9822
Ken-ichi Tadaki https://orcid.org/0000-0001-9728-8909
Rodrigo Herrera-Camus https://orcid.org/0000-0002-2775-0595
Gabriel Brammer https://orcid.org/0000-0003-2680-005X
Ivelina Momcheva https://orcid.org/0000-0003-1665-2073
Pieter van Dokkum https://orcid.org/0000-0002-8282-9888

References

Abramson, L. E., & Morishita, T. 2018, ApJ, 858, 40
Bigiel, F., Leroy, A., Walter, F., et al. 2008, AJ, 136, 2846
Bolatto, A. D., Warren, S. R., Leroy, A. K., et al. 2015, ApJ, 809, 175
Boselli, A., Fossati, M., Gavazzi, G., et al. 2015, A&A, 579, A102
Bouché, N., Dekel, A., Genzel, R., et al. 2010, ApJ, 718, 1001
Brammer, G. B., van Dokkum, P. G., Franx, M., et al. 2012, ApJS, 200, 13
Burkert, A., Förster Schreiber, N. M., Genzel, R., et al. 2016, ApJ, 826, 214
Calistro Rivera, G., Hodge, J. A., Smail, I., et al. 2018, ApJ, 863, 56
Calzetti, D., Armus, L., Bohlin, R. C., et al. 2000, ApJ, 533, 682
Chang, Y.-Y., van der Wel, A., Rix, H.-W., et al. 2013, ApJ, 773, 149
Chen, C.-C., Hodge, J. A., Smail, I., et al. 2017, ApJ, 846, 108
Conroy, C., Gunn, J. E., & White, M. 2009, ApJ, 699, 486
Dalcanton, J. J., Spergel, D. N., & Summers, F. J. 1997, ApJ, 482, 659
Davies, R. I. 2007, MNRAS, 375, 1099
Dekel, A., & Burkert, A. 2014, MNRAS, 438, 1870
Dutton, A. A. 2009, MNRAS, 396, 121
Erwin, P. 2015, ApJ, 799, 226
Erwin, P., Pohlen, M., & Beckman, J. E. 2008, AJ, 135, 20
Fakhouri, O., Ma, C.-P., & Boylan-Kolchin, M. 2010, MNRAS, 406, 2267
Fall, S. M., & Romanowsky, A. J. 2018, ApJ, 868, 133
Förster Schreiber, N. M., Genzel, R., Lehnert, M. D., et al. 2006, ApJ, 645, 1062
Förster Schreiber, N. M., Genzel, R., Bouché, N., et al. 2009, ApJ, 706, 1364
Förster Schreiber, N. M., Renzini, A., Mancini, C., et al. 2018, ApJS, 238, 21
Fossati, M., Fumagalli, M., Boselli, A., et al. 2016, MNRAS, 455, 2028
Fossati, M., Gavazzi, G., Savorgnan, G., et al. 2013, A&A, 553, A91
Fossati, M., Wilman, D. J., Mendel, J. T., et al. 2017, ApJ, 835, 153
Fumagalli, M., Fossati, M., Hau, G. K. T., et al. 2014, MNRAS, 445, 4335
Gavazzi, G., Consolandi, G., Dotti, M., et al. 2015, A&A, 580, A116
Genzel, R., Tacconi, L. J., Eisenhauer, F., et al. 2006, Natur, 442, 786
González Delgado, R. M., Cid Fernandes, R., Pérez, E., et al. 2016, A&A, 590, A44
Grogin, N. A., Kocevski, D. D., Faber, S. M., et al. 2011, ApJS, 197, 35
Hill, A. R., van der Wel, A., Franx, M., et al. 2019, ApJ, 871, 76
Hirschmann, M., Naab, T., Davé, R., et al. 2013, MNRAS, 436, 2929
Kassin, S. A., Weiner, B. J., Faber, S. M., et al. 2012, ApJ, 758, 106
Kelly, B. C. 2007, ApJ, 665, 1489
Kelvin, L. S., Driver, S. P., Robotham, A. S. G., et al. 2012, MNRAS, 421, 1007
Kennicutt, R. C., Jr. 1998, ARA&A, 36, 189
Koekemoer, A. M., Faber, S. M., Ferguson, H. C., et al. 2011, ApJS, 197, 36
Lang, P., Förster Schreiber, N. M., Genzel, R., et al. 2017, ApJ, 840, 92
Lang, P., Wuyts, S., Somerville, R. S., et al. 2014, ApJ, 788, 11
Lelli, F., McGaugh, S. S., & Schombert, J. M. 2016, ApJL, 816, L14
Li, H., Wuyts, S., Lei, H., et al. 2019, ApJ, 872, 63
Lilly, S. J., & Carollo, C. M. 2016, ApJ, 833, 1
Lilly, S. J., Carollo, C. M., Pipino, A., Renzini, A., & Peng, Y. 2013, ApJ, 772, 119
Livermore, R. C., Jones, T. A., Richard, J., et al. 2015, MNRAS, 450, 1812
Madau, P., & Dickinson, M. 2014, ARA&A, 52, 415
Marinacci, F., Fraternali, F., Nipoti, C., et al. 2011, MNRAS, 415, 1534
Markwardt, C. B. 2009, in ASP Conf. Ser. 411, Astronomical Data Analysis Software and Systems XVIII, ed. D. A. Bohlender, D. Durand, & P. Dowler (San Francisco, CA: ASP), 251
McGrath, E. J., Stockton, A., Canalizo, G., Iye, M., & Maihara, T. 2008, ApJ, 682, 303
Mo, H. J., Mao, S., & White, S. D. M. 1998, MNRAS, 295, 319
Momcheva, I. G., Brammer, G. B., van Dokkum, P. G., et al. 2016, ApJS, 225, 27
Nelson, E. J., van Dokkum, P. G., Brammer, G., et al. 2012, ApJL, 747, L28
Nelson, E. J., van Dokkum, P. G., Förster Schreiber, N. M., et al. 2016a, ApJ, 828, 27
Nelson, E. J., van Dokkum, P. G., Momcheva, I. G., et al. 2016b, ApJL, 817, L9
Nelson, E. J., Tadaki, K., Tacconi, L. J., et al. 2019, ApJ, 870, 130
Newman, A. B., Belli, S., & Ellis, R. S. 2015, ApJL, 813, L7
Noeske, K. G., Weiner, B. J., Faber, S. M., et al. 2007, ApJL, 660, L43

Pastrav, B. A., Popescu, C. C., Tuffs, R. J., & Sansom, A. E. 2013, A&A, 557, A137
Peng, C. Y., Ho, L. C., Impey, C. D., & Rix, H.-W. 2010, AJ, 139, 2097
Pezzulli, G., Fraternali, F., & Binney, J. 2017, MNRAS, 467, 311
Saintonge, A., Kauffmann, G., Wang, J., et al. 2011, MNRAS, 415, 61
Schreiber, C., Pannella, M., Elbaz, D., et al. 2015, A&A, 575, A74
Sérsic, J. L. 1968, Atlas de Galaxias Australes (Córdoba: Obs. Astron., Univ. Nac. Córdoba)
Sharples, R., Bender, R., Agudo Berbel, A., et al. 2012, Proc. SPIE, 8446, 84460K
Skelton, R. E., Whitaker, K. E., Momcheva, I. G., et al. 2014, ApJS, 214, 24
Storey, P. J., & Zeippen, C. J. 2000, MNRAS, 312, 813
Stott, J. P., Swinbank, A. M., Johnson, H. L., et al. 2016, MNRAS, 457, 1888
Struck, C., & Elmegreen, B. G. 2017, MNRAS, 469, 1157
Suess, K. A., Kriek, M., Price, S. H., & Barro, G. 2019, ApJ, 877, 103
Tacconi, L. J., Genzel, R., Saintonage, A., et al. 2018, ApJ, 853, 179
Tacconi, L. J., Neri, R., Genzel, R., et al. 2013, ApJ, 768, 74
Tadaki, K.-i., Genzel, R., Kodama, T., et al. 2017, ApJ, 834, 135
Toft, S., Zabl, J., Richard, J., et al. 2017, Natur, 546, 510
Turner, O. J., Cirasuolo, M., Harrison, C. M., et al. 2017, MNRAS, 471, 1280
van den Bosch, F. C. 2001, MNRAS, 327, 1334
van der Wel, A., Chang, Y.-Y., Bell, E. F., et al. 2014b, ApJL, 792, L6
van der Wel, A., Franx, M., van Dokkum, P. G., et al. 2014a, ApJ, 788, 28
van der Wel, A., Rix, H.-W., Wuyts, S., et al. 2011, ApJ, 730, 38
van Dokkum, P. G., Brammer, G., Fumagalli, M., et al. 2011, ApJL, 743, L15
van Dokkum, P. G., Leja, J., Nelson, E. J., et al. 2013, ApJL, 771, L35
van Dokkum, P. G., Nelson, E. J., Franx, M., et al. 2015, ApJ, 813, 23
Whitaker, K. E., Franx, M., Leja, J., et al. 2014, ApJ, 795, 104
Williams, R. J., Quadri, R. F., Franx, M., van Dokkum, P., & Labbé, I. 2009, ApJ, 691, 1879
Wisnioski, E., Förster Schreiber, N. M., Fossati, M., et al. 2019, ApJ, 886, 124
Wisnioski, E., Förster Schreiber, N. M., Wuyts, S., et al. 2015, ApJ, 799, 209
Wisnioski, E., Mendel, J. T., Förster Schreiber, N. M., et al. 2018, ApJ, 855, 97
Wuyts, E., Wisnioski, E., Fossati, M., et al. 2016, ApJ, 827, 74
Wuyts, S., Förster Schreiber, N. M., van der Wel, A., et al. 2011, ApJ, 742, 96
Wuyts, S., Förster Schreiber, N. M., Genzel, R., et al. 2012, ApJ, 753, 114
Wuyts, S., Förster Schreiber, N. M., Nelson, E. J., et al. 2013, ApJ, 779, 135